BLS WORKING PAPERS

U.S. DEPARTMENT OF LABOR
Bureau of Labor Statistics

OFFICE OF PRODUCTIVITY AND TECHNOLOGY

On the Estimation of Classical Human Capital Wage Equations with Two Independent Sources of Data on Actual Work Experience

Linda L. Moeller, U.S. Bureau of Labor Statistics

Working Paper 362
June 2002

On the Estimation of Classical Human Capital Wage Equations

With Two Independent Sources of Data on Actual Work Experience

Linda L. Moeller
Division of Productivity Research
Office of Productivity and Technology
Bureau of Labor Statistics

Revised May 28, 2002
Original June 4, 1998

This research was conducted under an inter-agency agreement between the Bureau of Labor Statistics and the Bureau of the Census. The research results and conclusions discussed here are those of the author and do not reflect the views of the Department of Labor, the Bureau of Labor Statistics, the Department of Commerce or the Bureau of the Census.

Abstract: The BLS multifactor productivity series decomposes labor productivity growth into components associated with increased capital intensity of production and shifts in the skill-composition of the work force attributable to changes in the level and distribution of human capital. The use of administrative record data on actual accumulated work experience as an indicator of workers' current productivity is a distinguishing features of this series. The current procedure relies on a one-time match of Social Security Administration (SSA) data to records from the March 1973 CPS to develop an experience proxy that is entered in a human capital wage equation. Since the parametric relationship between accumulated work experience and the demographic characteristics of the work force is unlikely to remain stable over time, the BLS has undertaken a long-run research project to update the work experience data at regular intervals.

Two independent sources of information on accumulated work experience were examined: SSA administrative record data on quarters of covered employment, and data from the 1984 Survey of Income and Program Participation (SIPP) on the number of years in which the respondent was employed for 6 months or longer. Direct comparison of the SIPP and SSA data show that the SSA data generate biased estimates of total actual work experience due to incomplete program coverage at the beginning of the Social Security program, in 1937, and subsequent changes in SSA coverage rates. Contemporaneous experience profiles estimated with SSA data are flatter than their SIPP counterparts, while the 1973 CPS-SSA experience profiles are flatter than their 1984 counterparts. Wage profiles estimated with the 1973 CPS-SSA experience proxy are too high for younger and older workers. Consequently weights assigned to the hours growth rates for younger and older workers are too large, while those assigned to hours growth rates for prime-aged workers are too small. These results tell a cautionary tale with respect the assumption that the coefficients of the experience equation are stable over time.

On the Estimation of Classical Human Capital Wage Equations

With Two Independent Sources of Data on Actual Work Experience

Linda L. Moeller

Table of Contents

Introduction

This paper discusses results from a research project in which the ultimate objective is the construction and analysis of a new BLS index of the skill-composition of the work force with annual data on the accumulated employment experience of U.S. workers. [1] Within the BLS multifactor productivity series, the labor composition index and the capital services index respectively serve as measures of the contributions of human and physical capital services to U.S. productivity growth.[2] The labor composition index is motivated by the widely-accepted view that workers become more efficient as they accumulate education and market work experience, and that economy-wide increases in average levels of education and work experience contribute to observed growth in productivity. Within the conceptual framework of models of long-run economic growth the labor composition index converts growth in hours worked into growth in labor efficiency units.[3]

---

[1] I would like to thank Judy Eargle, Enrique Lamas, and the entire staff of the Housing and Household Economic Statistics Division of the Census Bureau (HHES) for their generous hospitality during this project. This research has also benefited from the advice and encouragement of Phoebus J. Dhrymes, Howard M. Iams, David B. McMillen, Larry Rosenblum and Harold W. Watts. Any errors that remain are my responsibility. The results reported here pertain exclusively to the 1984 panel of the SIPP. Results from the period 1984-1993 are discussed in Linda Moeller (1999), "A Second Decade of Productivity Growth: Prototype Labor Composition Indexes Based on the SIPP," presented at the 1999 NBER Summer Institute Workshop on Price Output and Quality Measurement.

I am particularly indebted to Ms. Stacy Masano of the HHES staff, who assembled several SAS data sets containing information from selected SIPP topical modules and a subset of the variables available in the SIPP-SSA matched file for this project. Her careful work greatly facilitated my initital progress.

[2] The procedures currently used by BLS, as described in BLS Bulletin 2426, Labor Composition and Productivity Growth, 1948-90, U. S. Department of Labor, Bureau of Labor Statistics, December 1993. This paper describes possible improvements to the current approach, focusing on survey and administrative record data from 1984.

[3] Many prominent studies of the sources of productivity growth, such as those by Jorgenson, Gollop and Fraumeni, also contain estimates of the contribution of changes in the composition of the workforce. For example, see Dale W. Jorgenson, Frank M. Gollop and Barbara M. Fraumeni (1987), Productivity and U.S. Economic Growth, Cambridge, Mass.: Harvard University Press.

The BLS labor composition index is a hedonic index in which education and experience levels are assumed to be fundamental structural determinants of workers' productivity and market wage rates.[4] Regression coefficients from a standard human capital wage equation are used to estimate within-cell conditional mean wage rates for workers with different levels of experience and education. These conditional mean wage rates are used to estimate the labor cost shares attributable to workers with different sets of skill characteristics. The labor cost shares are then used to aggregate rates of growth of hours worked across cells. These calculations are described in more detail below.

The use of administrative record data on actual acccumulated work experience is one of the distinguishing characteristics of the current BLS labor composition index. Actual work experience is a particularly important explanatory variable in the determination of female wage rates because the labor force participation of women can be intermittent. Labor economists have appreciated the importance of long-run increases in the labor force participation rates of women to growth in market-oriented production, perhaps one of the most dramatic reorganizations of productive activity observed during this century, at least since the seminal work of Mincer, Becker, and Gronau. A substantial portion of this "first generation" labor supply literature was devoted to the estimation of labor supply and earnings equations with relatively large, nationally representative cross-section sets of microdata. The standard data base for many of these efforts

[4] This general approach is discussed extensively in Zvi Griliches (1970), "Notes on the Role of Education in Production Functions and Growth Accounting," in Education, Income and Capital, W. Lee Hanson, Ed., New York: NBER and Columbia University Press, pp. 71-115; and in Griliches (1971), "Introduction: Hedonic Price Indexes Revisited," in Price Indexes and Quality Change: Studies in New Methods of Measurement, Griliches, Ed., Cambridge: Harvard University Press, pp. 3-15. For a contemporary critical discussion of this approach see Ariel Pakes (2002), "A Reconsideration of Hedonic Price Indices with an Application to PC's," NBER Working Paper No. 8715. Some of Pakes' points also arise implicitly in this paper and in Moeller (1999), cited above, to the extent that I discuss the problem of selection bias as it arises in the estimation of wage equations and the roles of both supply and demand factors in the determination of market wage rates. But for the most part this paper follows a conventional hedonic approach.

has been the Current Population Survey (CPS), at least in part because its careful and consistent design have supported regular efforts to replicate significant findings. Subsequently several influential nationally representative surveys, notably the National Longitudinal Surveys (NLS) and the Panel Study of Income Dynamics (PSID), spawned increasingly sophisticated methodological work to deal with problems of heterogeneity and change over time that could not be addressed with the CPS.

From time to time, important complementary research has been conducted with large sets of confidential administrative record data.[5] Administrative record data from the Social Security Administration or the Internal Revenue Service have been matched to CPS microdata on a number of occasions in order to construct data sets that could be used for analytical purposes. In particular, experience equation coefficients estimated with microdata from the March 1973 CPS matched to Social Security Administration (SSA) data are used to construct a proxy for actual work experience in wage equations that are estimated annually by the BLS Office of Productivity and Technology (OPT). But in view of the recognized secular increase in women's labor force participation rates mentioned above, it seems unlikely that the true parametric relationship between actual work experience and the explanatory variables of the experience equation has been stable over time. Consequently BLS has undertaken a long-run project to update its labor composition measures.

This paper focuses on results obtained during the first stage of this project, in which actual work experience data from the 1984 panel of the Survey of Income and Program Participation (SIPP) were compared with actual work experience data from SSA administrative records, and wage

---

[5] Insightful prior research with survey data linked to SSA administrative records is illustrated in Nancy Ruggles and Richard Ruggles (1977), "The Anatomy of Earnings Behavior," The Distribution of Economic Well-Being, New York: NBER, pp. 115-158.

equations that are intermediate products in the composition of BLS's labor composition index were estimated with work experience data from each of these two independent data sources. Two shortcomings of the data with which BLS currently measures changes in the efficiency of labor services, and two shortcomings in the estimation procedures with which the data are aggregated into an index of shifts in the skill-composition of hours worked are discussed in this paper. These shortcomings have been recognized by BLS for a number of years; this research reflects a substantial effort to address them.[6]

The first shortcoming of current data sources is that estimates of annual earnings and annual hours worked are based on responses to questions from the annual March supplement to the CPS that pertain to the longest job held in the prior year. Many survey respondents hold more than one job during the course of a year, and many hold two jobs simultaneously. In a sense, the current estimates of the total number of earnings and hours worked during the course of the prior year are somewhat imprecise. The wage equation parameter estimates reported below have been obtained with microdata from the SIPP, an important and relatively new source of nationally representative panel data on earnings and hours worked that does not have these limitations.

The second shortcoming of current data sources is that information on actual accumulated work experience, which is an important determinant of average wage rates from the perspective of human capital theory, is not available on an annual basis. The current procedure relies on the estimated coefficients of a single set of work experience equations that have been estimated with a data set constructed by matching individual earnings records from the SSA to microdata from the March 1973 CPS, as noted above. Results from a direct comparison of 1984 SIPP estimates

---

[6] In particular, see the BLS Handbook of Methods (1997), BLS Bulletin 2490, pp. 90-92, and Labor Composition and Productivity Growth, 1948-90 (1993), BLS Bulletin 2426, Appendix G.

of total work experience to similarly linked SSA administrative records reveals that the latter are biased downward due to incomplete program coverage at the beginning of the Social Security program, in 1937, and subsequent changes in SSA coverage rates over time.[7] That is, the work experience recorded in the administrative records is known to be an underestimate of true work experience.[8]

The first methodological shortcoming discussed below is that the current procedure involves the estimation of a conventional wage equation, but does not incorporate a selection bias correction factor. Although it was argued in the early labor supply literature that the estimation of wage equations without accounting for the probability of employment resulted in coefficient estimates that were consistent "for the population of interest," over time the incorporation of an additive selection bias correction factor into the procedures with which wage equations are estimated has become more standard practice. Incorporation of a conventional selection bias correction factor that follows an early specification of Heckman is found to have small, but economically meaningful effects on the estimated earnings profiles of women, especially among highly-educated women.

[7] The SIPP was designed to permit matching with administrative records, in order to facilitate the analysis of issues related to the administration of income transfer and other Federal programs that cannot be accomplished with administrative record data alone. The files analyzed in the research reported here were originally linked for an SSA research project that is described in Howard M. Iams (1991), "Child Care Effects on Social Security Benefits," 1991 Annual Research Conference Proceedings, Bureau of the Census, pp. 255-271.

[8] This bias is worthy of some note, in light of a recent upsurge of interest in the cost savings that would result from increased use of administrative record data to construct Federal statistics. Several sessions of the Census Bureau's 1995 Annual Research Conference were devoted to this topic. In particular, see the following papers in the forthcoming Conference Proceedings: Ib Thompson, "Use of Administrative Records in the Norwegian Censuses Since 1970;" Pekka Myrskyla, Joe Knott and Cynthia Taeuber, "Uses of Administrative Records in Censuses and Health Research: Finland and the United States," and Ann Brown and Richard Veevers, "The Use and Evaluation of Administrative Records in the Self-Sufficiency Project." Some of the limitations of administrative record data for analytical purposes are emphasized in Janet L. Norwood (1985), "Administrative Statistics: A BLS Perspective," Journal of Business and Economic Statistics, 3(4), pp. 398-400.

The second shortcoming of the current methodology is that it implicitly imposes a number of structural restrictions on the estimated parameters of the wage equation.[9] It is shown below that these implicit restrictions distort the estimated wage profiles of older workers, presumably because the true structural relationship among the explanatory variables that are common to both the experience equation and the wage has not remained stable over time. The coefficient of the squared work experience proxy, which would otherwise cause estimated wage profiles to increase at a decreasing rate, becomes very small. Consequently estimated wage profiles become almost linear. Since the labor composition index cost-share weights are constructed with predicted wage rate values, the result is that the weights associated with growth in hours worked by younger and older workers are too large, and the weights associated with growth in hours worked by prime-aged workers too small under the current approach. This problem does not arise when contemporaneous work experience data are employed to estimate the experience equation.

Interestingly, earnings profiles generated with the estimated parameters of an equation in which potential experience is a proxy for actual work experience are similar to those obtained with current-year SIPP- and SSA-based experience estimates. Since potential experience is a function of age and years of schooling completed, and schooling dummies also enter the earnings equation directly, this result may indicate that the potential experience variable is capturing cohort effects or other compositional effects that influence the estimated parameters of the earnings equation.[10]

---

[9] These restrictions can be shown by explicitly substituting the right-hand-side of the experience equation into the wage equation, and comparing the result with the wage equation that is actually estimated.

[10] These compositional, or aggregation effects are discussed, for example, in Ray C. Fair and Kathryn Dominguez (1991), "Effects of the Changing U.S. Age Distribution on Macroeconomic Equations," American Economic Review, pp. 1276-1294; and in Thomas M. Stoker (1986), "Simple Tests of Distributional Effects on Macroeconomic Equations," Journal of Political Economy, pp. 763-795.

Description of the Labor Composition Index

Consider the secular increases in stocks of assets relative to the number of employed persons, and in the labor force participation rates of males and females, that are portrayed in Charts 1 and 2. The rising trends in capital intensity and female labor force participation rates suggest that increasing fractions of successive cohorts of women have found their marginal productivity in market production to be higher than their marginal productivity in household or other non-market production.[11] The concurrent long-run decline in labor force participation rates among men is largely attributable to increasingly early retirement.[12] The trend toward earlier retirement among men may in turn be due to declines in the relative marginal productivity of older male workers, as labor force participation rates of prime-aged females increase. As noted above, this reallocation of labor between the market and non-market sectors and the growing capital intensity of production appear to be two of the fundamental economic forces that have contributed to long-run growth in the productivity of labor.[13]

---

[11] A formal discussion of the determinants of the family labor supply decision is provided in Ruben Gronau (1986), "Home Production - A Survey." In Handbook of Labor Economics, Vol. I, Ashenfelter and Layard, Eds., Amsterdam: North-Holland, pp. 274-304. Strong empirical evidence on the increased labor force participation of women is provided in Goldin's study of labor force participation rates and average years of work experience among women from the last decade of the 1800's to the present. Labor force participation rates among married women increased with age in every cohort examined and labor force participation rates among new entrants increased with each successive cohort, while average levels of work experience among employed women increased slowly. See especially Figure 1 in Claudia Goldin (1989), "Life-Cycle Labor Force Participation of Married Women: Historical Evidence and Implications," Journal of Labor Economics, 7(1), pp. 20-47. Also see Mark R. Killingsworth and James J. Heckman (1986), "Female Labor Supply: A Survey," in Handbook of Labor Economics, Volume I, O. Ashenfelter and R. Layard, Eds., pp. 103-204.

[12] John Penceval (1986), "Labor Supply of Men: A Survey," in Handbook of Labor Economics, Volume I, O. Ashenfelter and R. Layard, Eds., pp. 3-102.

[13] For example, Jorgenson and Fraumeni assert that "The contributions of capital and labor inputs are the predominant sources of U. S. economic growth for the period as a whole [1948-1986]," in Dale W. Jorgenson and Barbara Fraumeni (1992), "Investment in Education and U. S. Economic Growth," Scandinavian Journal of Economics, 94(0), Supplement, pp. 51-70. Also see William J. Baumol, Sue

A familiar abstract representation of the employment decision associated with this process is provided in Chart 3.[14] The distribution of wage offers for a given category of labor at time t is represented by the left-most distribution, drawn with a dashed line. The acceptance wage of a person with a particular vector of demographic and household characteristics is represented by the vertical line. A person's acceptance wage is assumed to be determined by the relative marginal productivities of all household members in market and non-market production, their time constraints, and non-labor income. Persons who receive wage offers to the right of the vertical line accept them, and persons who receive wage offers to the left of the vertical line undertake non-market production. The probability that the person will be employed in period t is represented by the total area under the dashed line and to the right of the vertical line.

If increased capital intensity raises the marginal productivity of labor at time t+1, all else equal, the wage offer distribution for this category of labor will shift to the right. For t+1, the wage offer distribution becomes the distribution drawn with a solid line. If the composition of the population is unchanged between t and t+1, the probability increases that a person with the same observable characteristics and the same reservation wage will receive a wage offer above his or her reservation wage. In Chart 3, the probability of employment for a person within this category of labor becomes the larger area under the solid line, and to the right of the vertical line. Labor force participation rates will increase for this category of labor, holding other effects constant.

Anne Batey Blackman, and Edward N. Wolff (1989), Productivity and American Leadership: The Long View, Cambridge: MIT Press.

[14] The characterization of the labor force participation decision that follows is a simplified version of one used in many characterizations of the labor supply decision. See, for example, Mark R. Killingsworth (1983), "Second-generation Studies of Static Labor Supply Models: Methodology and Empirical Results," Chapter 4 in Labor Supply, Cambridge: Cambridge University Press.

Notice that the observed mean wage rate depends on the labor force participation rate, because market wage rates are not distributed symmetrically about the mean wage offer.[15]

Of course, increased capital intensity of production is not the sole determinant of observed trends in real wage rates. Education is widely believed to be a determinant of productivity growth because it increases skill levels observed during subsequent employment, and/or because it refines native abilities. Chart 4 shows that hours-weighted averages of years of school completed for men and women have increased, and converged, during the post-war period. The wage premiums associated with increased levels of schooling have been well documented.[16] And wage premiums associated with seniority and total work experience have been attributed to increased productivity due to formal and informal on-the-job training, increases in efficiency that come with experience at performing a particular set of work tasks, and with improved knowledge of the organizational or institutional structures through which work is organized.[17]

Average levels of work experience are believed to have declined during the 1970's as unusually large cohorts of new entrants joined the labor force and female labor force participation rates continued to increase. This demographic shift may have contributed to the decline in productivity growth rates that was observed during the 1970's. Similarly during the 1980's, cohorts of new

---

[15] The use of a single category of labor to illustrate the problem of selection bias follows the lucid discussion in Penceval (1986), "Labor Supply of Men," cited above.

[16] For evidence on the wage premiums associated with schooling see, for example, Marvin H. Kosters (1991), "Wages and Demographics," and the following commentary by Gary Burtless, in Workers and Their Wages, Marvin H. Kosters, Ed., Washington: AEI, pp. 1-38. Theodore W. Schultz provides an excellent early discussion of the view that education contributes to productivity in his 1963 essay, The Economic Value of Education, New York: Columbia University Press.

[17] For example, Topel estimates that a 10-year increase in job seniority is associated with a 25 percent increase in the wage rate. See Robert Topel (1991), "Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority," Journal of Political Economy, pp. 145-176. Jacob Mincer provides a strong conceptual and empirical discussion of the relationship between earnings and work experience in his seminal work of 1974, Schooling, Experience and Earnings, New York: NBER and Columbia University Press; see especially Chapter 4 of that volume.

labor market entrants declined in size and productivity growth rates increased.[18] The labor composition index is an attempt to quantify the contribution of these changes in the composition of the work force on multifactor productivity growth rates.

As noted by Medoff and Abraham, "earnings which are associated with more labor market experience at a point in time" are not necessarily "explained by higher productivity at the same point in time."[19] In addition to prior work experience, the current allocation of time between training and other productive activities that may be undertaken on the job is widely assumed to be a primary determinant of current earnings.[20] If some fraction of workers' time is allocated to on-the-job training and the costs of that training are shared between employees and employers, then the earnings of workers who are investing in training may be lower than the earnings of those who are not, all else equal. However, the finiteness of human life constrains the frequency and duration of these investments as work experience increases.[21] Consequently, over the long run

---

[18] Discussions of these demographic trends are provided in Kosters (1991), "Wages and Demographics," cited above; John Bound and George Johnson (1991), "Wages in the United States During the 1980s and Beyond," in Workers and their Wages: Changing Patterns in the United States, M. H. Kosters, Ed., pp. 77-103, and in the comment on Bound and Johnson from Lawrence Katz , pp. 104-106. Also see Jacob Mincer (1991), "Human Capital, Technology, and the Wage Structure," cited above; and BLS Bulletin 2426, Labor Composition, cited above.

[19] James L. Medoff and Katherine G. Abraham (1981), "Are Those Paid More Really More Productive? The Case of Experience," Journal of Human Resources, 16, pp. 186-216. Also see Judith K. Hellerstein and David Neumark (1995), "Are Earnings Profiles Steeper than Productivity Profiles? Evidence from Israeli Firm-Level Data," Journal of Human Resources, pp. 89-112.

[20] For example, in their analysis of data on entry-level positions Barron, Black and Loewenstein find that a 10% increase in training is associated with a 1.5% increase in earnings, but a 3% increase in productivity. See John M. Barron, Dan A. Black, and Mark A. Loewenstein (1989), "Job Matching and On-the-Job Training," Journal of Labor Economics, 7, pp. 1-19. For early estimates of time-investment profiles when the share of time devoted to on-the-job training is an unobserved explanatory variable, see James J. Heckman (1975), "Estimates of a Human Capital Production Function Embedded in a Life-Cycle Model of Labor Supply, " in Household Production and Consumption, Nestor E. Terleckyj, Ed., Studies in Income and Wealth by the Conference for Research in Income and Wealth, volume 40, New York: NBER and Columbia University, pp. 227-264.

[21] A strong theoretical discussion of this point is provided in Alan S. Blinder and Yoram Weis (1976), "Human Capital and Labor Supply: A Synthesis," Journal of Political Economy, 84 (3), pp. 449-472.

both education and work experience remain important explanatory variables, for an analysis of the contribution of changes in the composition of hours worked to productivity within a national growth-accounting framework.

The labor composition index is the component of the BLS multifactor productivity series that is intended to reflect changes in the skill composition of the work force. The overall approach is to calculate multifactor productivity growth as the difference between the growth of output and a weighted sum of inputs, assuming that the inputs are paid the value of their marginal products, on average. Assuming that an aggregate production of the form $Q_t = A_t F\left(K_t, L_t\right)$ exists and that capital and labor services are paid the value of their marginal products, logarithmic derivatives of the production function can be rearranged to obtain the following expression for multifactor productivity growth, $\frac{\dot{A}_t}{A_t}$.[22]

$$\frac{\dot{A}_t}{A_t} = \frac{\dot{Q}_t}{Q_t} - s_K \frac{\dot{K}_t}{K_t} - s_L \frac{\dot{L}_t}{L_t},$$

where s_K, s_L denote the shares of capital and labor respectively in national income.

The labor composition index decomposes growth in labor services, or $\frac{\dot{L}_t}{L_t}$, into growth in hours worked and changes in the skill-composition of the work force. Labor composition index weights ω_l are defined to be two-year averages of the labor cost shares of workers with characteristics l, as discussed below. Letting $H_{l,t}$ denote hours worked by workers with characteristics l at time t, omitting subscripts from aggregated values, and omitting time subcripts for simplicity:

---

[22] This exposition is generally well known. See, for example, Charles R. Hulten (2001), "Total Factor Productivity: A Short Biography," in New Developments in Productivity Analysis, edited by Charles R. Hulten, Edwin R. Dean and Michael J. Harper, NBER and CRIW Studies in Income and Wealth, Vol. 63, Chicago: University of Chicago Press, pp. 1-53.

$$\frac{\dot{L}}{L} = \omega_1 \frac{\dot{H}_1}{H_1} + \cdots + \omega_n \frac{\dot{H}_n}{H_n} = \omega_1\left(\frac{\dot{H}_1}{H_1} - \frac{\dot{H}}{H}\right) + \cdots + \omega_1\left(\frac{\dot{H}_n}{H_n} - \frac{\dot{H}}{H}\right) + \frac{\dot{H}}{H} \equiv \frac{\dot{C}}{C} + \frac{\dot{H}}{H}.$$

Thus $\frac{\dot{C}}{C} \equiv \frac{\dot{L}}{L} - \frac{\dot{H}}{H}$. Just as multifactor productivity growth rates are an estimate of the degree to which labor productivity growth is attributable to shifts in the composition of physical assets employed in production, the labor composition index is an estimate of the degree to which labor productivity and multifactor productivity growth rates are attributable to shifts in the skill-composition of total hours worked.

In practice the labor composition index is obtained by taking the exponent of the expression on the right-hand side of the following equation.

$$\ln C_t - \ln C_{t-1} = \left[\omega_1\left(\ln H_{1,t} - \ln H_{1,t-1}\right) + \cdots + \omega_n\left(\ln H_{n,t} - \ln H_{n,t-1}\right)\right] - \left[\ln H_t - \ln H_{t-1}\right].$$

A sequence of values for the "quality" of labor services employed in the market is calculated by setting $C_0 = 1$ and "chaining" the following expression:

$$C_t = C_{t-1} \exp\left\{\left[\omega_1\left(\ln H_{1,t} - \ln H_{1,t-1}\right) + \cdots + \omega_n\left(\ln H_{n,t} - \ln H_{n,t-1}\right)\right] - \left[\ln H_t - \ln H_{t-1}\right]\right\}.$$

BLS uses a conventional human capital wage equation to estimate conditional mean wage rates, $\hat{w}_{l,t}$, which are used in turn to calculate labor cost share weights for persons with characteristics l at time t:

$$\omega_{l,t} = \frac{1}{2}\left[\left(\hat{w}_{l,t} H_{l,t} \Big/ \sum_{\forall j,t} \hat{w}_{j,t} H_{j,t}\right) + \left(\hat{w}_{l,t-1} H_{l,t-1} \Big/ \sum_{\forall j,t-1} \hat{w}_{j,t-1} H_{j,t-1}\right)\right].$$

The conditional mean wage rates $\hat{w}_{l,t}$ are calculated as follows:

$$\hat{w}_{l,t} = \hat{\beta}_0 + \hat{\beta}_{s,t}\bar{s}_{l,t} + \hat{\beta}_{e.t}\bar{e}_{l,t} + \hat{\beta}_{d,t}\bar{d}_{l,t} + \hat{\beta}_{r,t}\bar{r}_{l,t}\ .$$

In this last equation an overbar denotes a within-cell weighted sample mean. Categorical variables (dummy variables) for the number of years of school completed are represented as $s_{l,t}$, and $e_{l,t}$ represents years of actual work experience for persons with characteristics l. The demographic variables in the vector $d_{l,t}$ include ever-married status, Black or ethnicity, and full-time/part-time status. Categorical variables in the vector $r_{l,t}$ include region and a city size variable.

Because the distinction between actual experience and potential experience may be an important one, especially in the case of women, BLS enters a proxy for actual experience as an explanatory variable in the wage equation. In light of the cohort-specific increases in the labor force participation and actual work experience of women that have been well-documented by Goldin, actual work experience data from successive panels of nationally representative surveys are to be preferred to the current proxy, which is based on a single CPS cross-section matched to Social Security administrative records.[23]

[23] Econometric procedures that might be implemented with the SSA data, to compensate for selection bias due to relatively low Social Security coverage rates at the inception of the program and time-varying ceilings on recorded earnings, are discussed in Honig and Hanoch (1985), cited above. However the simplifying assumptions that would be required to adjust for both of these limitations simultaneously are non-trivial. Also see Goldin (1989), cited above.

Measurement of Hours Worked

With respect to the measurement of growth in hours worked, the SIPP has two noteworthy strengths. First, SIPP calendar year estimates are based on information collected during three or four separate interviews. Questions on the number of hours usually worked per week, and on the number of weeks worked in each month of the 4-month reference period, are asked for each of two jobs, and for each of two types of self-employment. In contrast, the generality of the questions on hours worked in the previous calendar year in the annual supplement to the March CPS suggests that hours worked on intermittent jobs and second jobs may be excluded from the associated CPS estimates of total hours worked. Thus the coverage and accuracy of SIPP estimates of the number of hours worked in a calendar year dominate may estimates based on the annual March supplement to the CPS.[24]

Second, results from a PSID validation study, originally conducted to gauge respondent recall error, suggest that respondent recall errors in periodic questions on the number of weekly hours usually worked may tend to be serially independent. Consequently the measurement error contained in calendar year estimates from several SIPP interviews repeated during the course of a single calendar year may be relatively small. These points are developed in more detail below.

---

[24] The sample size of the CPS is substantially larger than that of the SIPP, but the discrepancy between the two has been diminishing in recent years.

March Current Population Survey (CPS) Questions on Hours Worked

The hours growth rates currently used to construct the labor composition index are estimated from answers to questions included in the annual March supplement to the CPS. Specifically, the total number of hours that the respondent worked during the preceding calendar year is estimated as $A \times B,$ where A and B are responses to the following questions.

> A: During 19.. in how many weeks did … work even for a few hours? Include paid vacations and sick leave as work.

> B: In the weeks that … worked how many hours did … usually work per week?

These questions are clearly designed to characterize the general labor force behavior of the work force; they are less than ideal as a source of information about the number of hours worked in an accounting sense.[25]

A separate question on the total number of jobs held has been asked of one-fourth of CPS respondents since 1994.[26] Analysis of averages for calendar year 1995, based on these data, reveals that roughly 6% of the employed held a second, as well as a primary job. Given the very

---

[25] During her tenure as a BLS Research Fellow Nancy Mathiowetz, a faculty member in the Joint Program in Survey Methodology, worked on a project to improve the reliability of March CPS-based estimates of the number of hours worked during the previous calendar year. She focused on the development of a separate set of survey questions for respondents whose employment is intermittent or otherwise non-standard.

[26] A similar question was asked periodically in the May survey, over the period 1970-1992. See John F. Stintson, Jr. (1997), "New data on multiple jobholding available from the CPS," Monthly Labor Review, March issue, p. 7.

general nature of the questions listed above, it seems seem likely that information on the number of hours worked on intermittent second jobs will be omitted from the product of A and B. Similarly the industry and occupation codes assigned to respondents who were employed during the previous calendar year are based on the longest-held job. Especially in the case of persons who are employed part-time it seems clear that hours worked on the longest-held job do not necessarily correspond to the industry and occupation in which the largest number of hours were worked.[27]

The CPS data indicate that the probability of holding a secondary job is largely constant across quintiles of weekly earnings on the primary job: 6.4 percent of respondents in the lowest quintile of weekly earnings (($0-210) and 5.9 percent of respondents in the highest quintile ($731 and over) were multiple jobholders.[28] However, multiple jobholding rates increase with education, from approximately 3.3 percent among persons with less than a high school diploma to 9.4 percent among persons with a PhD.

The occupational categories that accounted for the largest numbers of multiple jobholders, and for the highest multiple jobholding rates, are reported in the following Table. Industries with high multiple jobholding rates include services (7.6%), especially educational services (10.2%), and public administrators (9.2%), especially justice, public order and safety workers (11.9%).

---

[27] For evidence that women's career paths incorporate a high degree of occupational mobility, see Mary Corcoran, Greg J. Duncan, and Michael Ponze (1983), "Work Experience and Wage Growth of Women Workers," in Five Thousand American Families - Patterns of Economic Progress: Analyses of the First Thirteen Years of the Panel Study of Income Dynamics, Greg J. Duncan and James N. Morgan, Eds., Ann Arbor: Institute for Social Research, University of Michigan, pp. 249-323.

[28] Thomas Amirault (1997), "Characteristics of multiple jobholders," Monthly Labor Review, March issue, pp. 9-15.

Occupations of Multiple Jobholders			
Largest numbers of multiple jobholders	Rate	Largest rates of multiple jobholding	Rate
Salesworkers, retail & personal services	6.1	Firefighting occupations	28.1
Supervisors & proprietors, sales occup's	5.1	Physicians assistants	23.4
Mechanics & repairers	5.1	Announcers	19.3
Motor vehicle operators	5.4	Artists, performers, & related n.e.c.	16.0
Construction trades	4.2	Psychologists	15.6
Secretaries	6.0	Therapists	14.5
Writers, artists, entertainers, & athletes	9.3	Dental hygenists	14.4
Registered nurses	9.6	Teachers, college & university	14.1
Teachers, secondary school	13.3	Teachers, secondary school	13.3
Teachers, elementary school	8.8	Musicians & composers	13.0
Source: Amirault (1997), "Characteristics," Monthly Labor Review, Tables 4 and 5.			

Cross-tabulations of occupational and industry categories of the primary and secondary job show that secondary jobs are distributed across all categories, although a secondary job in the same industry or occupation as the primary job is observed most frequently. In particular, 53% of persons whose primary job is in the professional specialty occupations also hold secondary jobs in professional specialty occupations. Similarly 45 percent of secondary jobholders whose primary job was in services except private household services also held secondary jobs in services. Among workers in sales who also held secondary jobs, 23% held secondary jobs in sales.

Among secondary jobholders whose primary job was in occupations that may be characterized by cyclical employment variability, a substantial fraction of secondary jobs were also in occupations that include a significant cyclical component. In particular, the table below shows that roughly half of workers whose primary jobs were in sales, administrative support, or blue-collar work also held secondary jobs one of in the occupational categories also listed in the table. This pattern suggests that a significant percentage of secondary jobs may be demand-driven, with employment and hours worked on both the primary job and the secondary job increasing during cyclical upswings, and decreasing in downturns. To the extent that this is so, and to the extent that hours worked on these secondary jobs are omitted from the survey questions listed above, cyclical variability in hours worked may be underestimated in the CPS.

Percentage of secondary jobs in an occupation listed below, by occupation in primary job.	
Occupation in the primary job:	%
Sales	49.4
Administrative support	58.2
Precision production, craft and repair	51.5
Machine operators, assemblers, and inspectors	43.7
Transportation and material moving	48.8
Handlers, equipment handlers, and laborers	49.5
Source: Amirault, "Characteristics," Table 7.	

Prior Research on Respondent Recall Error

Results from validation studies, in which respondents' answers to questions about their earnings and hours worked are compared with establishment payroll records or Social Security Administration unemployment insurance records have been summarized by Rodgers, Brown and Duncan, and by Bound and Kreuger.[29] Particularly noteworthy is a comparison of household survey responses matched to payroll data from a large manufacturing establishment in which Rodgers, Brown and Duncan found that respondent recall errors in periodic questions on the number of weekly hours worked were serially uncorrelated. This result suggests that overall estimates of hours worked obtained by aggregating responses to these periodic questions were unbiased and, by extension, that respondent recall errors in successive SIPP interviews may also be uncorrelated, and largely offsetting, in calendar-year estimates of total hours worked.[30]

[29] Willard L. Rodgers, Charles Brown and Greg J. Duncan (1993), "Errors in Survey Reports of Earnings, Hours Worked, and Hourly Wages," Journal of the American Statistical Association, 88(424), December, pp. 1208-1218; and John Bound and Alan B. Kreuger (1991), "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" Journal of Labor Economics, 9 (1), pp. 1-24.

[30] However, Rodgers, Brown and Duncan found that contemporaneous respondent recall errors for questions on earnings and hours worked, asked within a single interview, were significantly positively

Measurement of Actual of Work Experience

The data examined during the initial stage of the project were extracted from longitudinal and topical module files of the 1984 Panel of the SIPP, and matched to administrative record data from the Social Security Administration (SSA). First, the frequency with which discrepancies arise between work experience estimates obtained with the two independent data sources examined (SIPP and SSA), the magnitudes of these discrepancies, and institutional reasons for them were assessed. These comparisons help to establish the integrity of the SIPP and the SSA data, and to establish a quantitative relationship between SIPP- and SSA-based work experience measures that can be used to evaluate differences in magnitudes in the associated experience and wage equation parameter estimates.

Some discrepancies between the SIPP and the SSA work experience variables are expected because the SIPP identifies spells of employment whose duration is six months or longer, while the SSA records "quarters of coverage," i.e., quarters in which the individual has been paid at least $50 in covered wages.[31] Years of employment reported in the SIPP would be a subset of years of employment recorded in the SSA records if the SSA covered all employment. But both

---

correlated. Like Bound and Kreuger, they found that respondent recall errors on earnings questions were significantly positively correlated in successive interviews.

[31] The full definition of the term "quarters of coverage" prior to 1978 was as follows.

> Calendar quarters in which a worker is paid $50 or more in wages for employment covered under the law (except wages for agricultural labor) or calendar quarters deemed to be quarters of coverage because the worker (1) was paid the maximum annual taxable earnings in a calendar year, (2) had self-employment net earnings of at least $400 in a taxable year, or (3) met the condition, defined by law, for acquiring a quarter of coverage through agricultural labor.

Social Security Bulletin, Annual Statistical Supplement, 1974, p. 12. Beginning in 1978, a quarter of coverage was credited for each $250 of annual earnings, with fractions truncated and earnings measured in 1978 dollars. Thus the SSA and SIPP data for the period 1978-1984 are less directly comparable than they are for earlier years. I thank Howard Iams for clarifying this point.

compulsory and elective coverage of employment by SSA has expanded significantly since SSA legislation was first enacted.

Neither the SSA nor the SIPP record industry of prior employment, so it is not possible to identify specific records of persons whose prior work history includes spells of employment not covered by SSA. Judging from population coverage rates, however, it is expected that roughly two-fifths of the early experience reported in the 1984 SIPP panel by persons who began working during or before the 1940s is not included in the SSA administrative record data, although virtually all of the employment in calendar year 1984 should be confirmed in SSA records.[32] This institutional shift in SSA coverage imparts a systematic downward bias to SSA-based estimates of work experience among mature workers. In fact SSA estimates of total work experience are smaller than SIPP estimates of total work experience in some cases, even though the SIPP collects little information on short, discrete spells of employment that occurred during the early stages of respondents' careers.

The SIPP work experience variables are subject to error because the recall period is the respondent's entire working life to date.[33] SSA data cannot be used to evaluate the degree of

---

[32] For example, in 1940 roughly 58% of all employed persons were covered by SSA, but by 1974 the coverage rate had increased to 90%. In 1980 the SSA coverage rate was 97%, and by 1985 it was 99%. Social Security Bulletin, Annual Statistical Supplement, 1974, Table 33, and Social Security Bulletin, Annual Statistical Supplement, 1991, Table 3.B.

[33] The likelihood that retrospective recall data on work experience may suffer from substantial measurement error was stressed by Greg J. Duncan during a helpful phone conversation in the beginning stages of this project. See Greg J. Duncan and Daniel H. Hill (1985), "An Investigation of the Extent and Consequences of Measurement Error," cited above; Nancy Mathiowetz and Greg Duncan (1988). "Out of Work, Out of Mind: Response Errors in Retrospective Reports of Unemployment." Journal of Business and Economic Statistics, pp. 221-229; John Bound, Charles Brown, Greg J. Duncan and Willard Rodgers (1989), "Measurement Error in Cross-Sectional and Longitudinal Labor Market Surveys: Results From Two Validation Studies," NBER Working Paper No. 2884; Robert Topel (1991), "Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority," Journal of Political Economy, 99, pp. 145-176; and James N. Brown and Audrey Light (1992), "Interpreting Panel Data on Job Tenure," cited above. Note, however, that the design of the SIPP and the design of the PSID are not directly comparable with respect to

respondent recall error in the SIPP with very much precision, because non-coverage of some industries and occupations, and shifts in coverage over time, mean that the incidence and duration of some true employment spells is not recorded in SSA records. Nonetheless the sensitivity of experience and wage equation parameter estimates to the use of alternative work experience measures can be examined, and the published time series of coverage rates provide a rough gauge of the incidence of annual discrepancies attributable to changes in SSA coverage for the population as a whole. Therefore a "broad brush" evaluation of the two variables is possible.

Internal discrepancies between alternative work experience measures within each data source were calculated, to provide a quantitative frame of reference within which to evaluate discrepancies that arise between the two independent data sources.[34] In the case of the SIPP, responses to a direct question on the total number of years worked since the first job lasting 6 months or longer were compared to estimates that have been constructed by summing the durations of all distinct employment spells reported.[35] In the case of the SSA, counts of years in which individuals worked 2 or more quarters were compared to counts of years in which the level of real annual earnings suggests that the respondent worked at least six months during the year.[36]

---

the collection of work experience data, and that Brown and Light find substantially less inconsistency in National Longitudinal Survey (NLS) work experience data than they do in the PSID.

[34] These comparisons of alternative measures from a single source also assist in debugging computer programs.

[35] A maximum of 6 discontinuous employment spells could be reported in responses to work history questions in the 1984 Panel of the SIPP .

[36] The focus on employment spells of 2 quarters or longer increases the comparability of SIPP and SSA measures of work experience. In prior work by OPT with a file in which microdata from the 1973 March CPS are matched to SSA administrative records, work experience is measured as the sum of all recorded quarters of employment.

Detailed Description of SIPP Work Experience Measures

Detailed questions on the duration of employment spells within respondents' recent and early work experience and a question on the total number of years worked are asked in a single SIPP interview.[37] Consequently checks on the internal consistency of the SIPP work experience data are limited to comparisons between the reported total number of years worked, on the one hand, and the total number of years in employment that may be inferred from responses to detailed early and recent work history questions, on the other hand.[38] The full set of detailed early and recent work history questions is asked only of currently or recently employed persons aged 21-65 who have not worked continuously since the first year in which they were employed for 6 months or longer. Persons with no spells out of employment lasting 6 months or longer are only asked the year in which they first worked 6 months or longer.

In the case of respondents between the ages of 21 and 65 who have not been employed continuously since their first job, questions are asked about beginning and ending dates for the first four employment spells lasting 6 months or longer.[39] In the case of respondents aged 16 to 20 older who are currently employed, or who have been employed during the past two-year period, information is only collected on the most recent job held. In the case of currently or recently employed persons aged 21 or older who held their current job for less than 10 years, information is also collected on the duration of a prior job.

[37] Questions about the respondents' current labor force status are also asked in each of 8 successive interviews, but since these interviews span a relatively brief period of 32 months they cannot be used to measure total work experience.

[38] Presumably internal inconsistencies in the SIPP will be reduced significantly with the implementation of computer-assisted interviewing procedures.

Distributions of years worked as reported directly by all respondents to the third topical module of the 1984 SIPP, in which the work history data are collected, are shown in Table 1 below. These distributions are reported separately by gender.[40] The distributions on the left of each table characterize all respondents, including persons who have been unemployed or out of the labor force for more than a year. The distributions on the right are for persons who are currently employed.[41] The availability of work experience variables for persons aged 16-21 is reported separately because, as noted above, the summary question on the total number of years worked and the detailed early work history questions are not asked of persons in this age group. Young adults are asked only about the duration of a current and prior job.

As expected, Table 1 shows that a larger fraction of the female population has fewer than 10 years of work experience, compared with men. A larger fraction of the male population has worked for longer than twenty years. It is interesting to note that roughly a fifth of the population of each gender reported 10 to 20 years of work experience in 1984. But the essential point for current purposes is that work experience data are available for SIPP sample observations that represent all but 1-2% of the employed population.

Detailed work history data are pertinent for the estimation of earnings and experience equations in which tenure on the current job is a separate explanatory variable from previously accumulated

[39] These questions do not distinguish among jobs that may have been held sequentially or concurrently within a single spell of employment.

[40] Total work experience has been calculated for persons who report having been employed continuously, because a separate question on the total number of years worked is not asked of these respondents.

[41] To maintain comparability with frequency distributions from the screened SSA series, 6050 observations (unweighted) on persons who report a total work experience of less than 6 months in the SIPP are counted as having 0 work experience. The vast majority (93%) of these observations are from persons who are not currently employed; 74% are from women over the age of 65. Among the employed, 58% of the 405 observations on persons who worked less than one year are from women.

work experience.[42] The availability of detailed data on the duration of employment spells is shown in Table 2, where total work experience estimates have been obtained by summing the employment spells of persons who have not been employed continuously since the first job that lasted 6 months or longer. Table 2 shows that SIPP work history data permit a distinction between tenure on the current job (or the last job held) and prior work experience for approximately 83% of the total population, and for 89% of the currently employed.[43]

---

[42] Mincer and Jovanovic find tenure on the current job to be a significant separate explanatory variable in wage equations for men, and indicate that this variable may capture returns to job-specific human capital. In work with the PSID, Abraham and Farber find evidence of small but significant returns to seniority among white collar males, and returns to seniority that are insignificantly different from zero in the case of blue collar males. Topel also finds positive returns to seniority, as noted above.

Marshall and Zarkin note that the duration of tenure is a function of the difference between the observed offered wage growth rate and the unobserved reservation wage growth rate of the employee, and estimate a two-step switching model to control for selectivity bias attributable to under-representation of persons for whom the offer wage after a trial period of employment is lower than the reservation wage of the employee after the trial period of employment. They find that tenure on the current job is not a significant explanatory variable in the context of their switching model, although it is significant when they follow the estimation procedures employed by Jovanovic and Mincer. However, Marshall and Zarkin work exclusively with data from young men, for whom accumulated stocks of specific human capital may be small. I am unaware of other research in which these results have been sustained in applications with microdata from older men. See Jacob Mincer and Boyan Jovanovic (1981), "Labor Mobility and Wages," in Studies in Labor Markets, S. Rosen, Ed., Chicago: University of Chicago Press; Katherine G. Abraham and Henry S. Farber (1987), "Job Duration, Seniority, and Earnings," American Economic Review, pp. 278-297; Marshall and Zarkin (1987)," The Effect of Job Tenure on Wage Offers," cited above; and Robert Topel (1991), "Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority," cited above.

[43] More specifically, detailed work history data are available for observations representing about 28% of the overall population, and about 27% of the currently employed. Persons representing an additional 44% of the total population, and 61% of the currently employed, report that they were employed continuously since the first year that they held a job that lasted 6 months or more. In addition 11% of the observations from the overall population, and 1% of the observations from the employed, are from persons aged 65 or older. Information on the current job, or on the job held most recently, is available for the latter two categories of observations (the continuously employed and persons over the age of 65), although a detailed work history is not. Among the remaining observations, those from persons aged 16-21 account for about 10% of the population, and 9% of the employed. Information on the duration of the current or most recent job held by these younger persons is available, but the total number of years worked is not.

Distributions of discrepancies of more than three years between reported and counted spells of work experience, both recorded in the SIPP, are reported separately by gender in Table 3.[44] Both measures of total work experience are available in the SIPP only for observations represented by the first three rows of each panel. Detailed work history questions are not asked of persons who report having been employed continuously since their first job, nor are they asked of persons aged 65 or older. As noted above, early work history questions are not asked of young adults. However the findings reported in Table 3 indicate that discrepancies between the two work experience measures are small for the vast majority of observations on prime-aged employed persons (persons between the ages of 16 and 65) who report some discontinuous employment. As expected given the higher incidence of spells out of the labor force among women, discrepancies occur among a larger fraction of records for women than they do among records for men.[45]

Cell means for total reported work experience, for which distributions are given in Table 1, have been compared with cell means for total calculated work experience, for which distributions are given in Table 2. The latter variable is obtained by summing the number of years spanned by the separate employment spells that are recorded in respondents' detailed work histories. Gender-specific averages within 5-year age intervals, not reported here, are consistently larger for calculated total work experience than for total reported work experience, and these discrepancies increase in absolute value with age. Among respondents whose entire work history is not covered in the detailed work history questions, the average duration of the period during which detailed information is not available also increases with age. These patterns suggest that some spells out

---

[44] Discrepancies of three years are less may be attributable to the fact that specific months in which employment spells began or ended are not recorded. For example, an employment spell that began in January of 1970 and ended in December of 1971 is likely to be counted as two years in the *reported* number of years worked, but would only contribute one year to the *counted* number of years worked.

[45] Cases in which the counted number of years worked is not available are generally ones in which the sequence of employment spells contains inconsistencies. The specific problem that arises is usually an

of employment of significant duration are omitted from the detailed work history questions of some respondents, and that the measurement error associated with these omissions is correlated with age and potential experience. Therefore the direct response to the total work history question is taken to be the preferred measure of total work experience in the SIPP.

Detailed Description of SSA Work Experience Measures

The SSA records examined consist of all sample observations from the SIPP work history topical module that could be matched to SSA administrative record data; a match was achieved for 92% of the SIPP observations. The SSA data consist of individual annual time series data on earnings, number of quarters of covered employment, number of quarters of covered self-employment, and number of quarters of covered agricultural employment.[46]

The shares of SIPP observations from a given age group that could be matched to SSA administrative record data are given in Table 4, along with the share of total observations represented by each group. For example, records on persons aged 16-21 accounted for 14% of all records in the SIPP 1984 Wave 3 file. Within that age group, 13% of the records were not matched to SSA data, and 87% were matched.[47] The shifts in the distributions of matched and

---

inconsistent sequence of dates for the beginning and ending of a spell of employment. There were 39 such cases among observations on women, and 6 among observations on men (unweighted).

[46] Data on quarters in of employment extend back to 1937, when the Social Security program was begun. Topcoded earnings data are avilable beginning in 1951. Self-employment data are available beginning in 1953, and agricultural employment data begin in 1955. To increase the comparability of these series with the SIPP information on number of years worked 6 months or longer, the SSA employment series were screened to include only observations in which two or more quarters of employment were reported. The agricultural and self-employment fields flag whether or not this type of employment was reported during the year but do not record the number of quarters worked beginning in 1978, when there was a shift in the coding scheme. See footnote 31, above, for more detail on the 1978 coding changes.

[47] A non-match does not necessarily mean that the SIPP respondent does not have a record at the Social Security Administration. In some instances survey respondents refuse to provide the necessary matching information in order to safeguard their privacy. In other instances the information provided by survey

non-matched records by cohort, shown in the remaining rows of Table 4, indicate that eventually most people work in an industry that is covered by SSA. However, the true duration of total employment cannot be determined from the SSA records for reasons given above, and illustrated in the tables that follow.

The SSA data show a relatively smooth increase in the incidence of covered (i.e., recorded) employment over time. In an attempt to increase the comparability between SSA data on the number of quarters when a person earned at least $50 in wages, on the one hand, and SIPP data on spells of employment lasting 6 months or longer on the other hand, information on the median weekly earnings of part-time workers was used to calculate the annual earnings of respondents who worked part-time for 25 weeks, and earned one-fourth of the median weekly earnings for part-time workers of the respondent's race and gender.[48] Earnings below this level were interpreted as evidence that the persons' employment spell had been shorter than six months. After implementation of this screen for short employment spells roughly 17% of all 1984 SIPP respondents were found to have worked in jobs covered by SSA in 1955, while 42 of all 1984 SIPP respondents worked in jobs covered by SSA in 1984.[49] This gradual increase reflects both the current age composition of the 1984 SIPP panel and the change in SSA coverage rates noted above. Many SIPP respondents had not yet reached working age by 1955; younger SIPP respondents were not yet born. But the true percentage of 1984 SIPP respondents who were

---

respondents is insufficient to achieve a match. The second case often arises in the case of proxy respondents, i.e., when one family member responds to the survey on behalf of another family member who is absent.

[48] This information is published in recent issues of Employment and Earnings ; topcoded earnings information is available in the SSA records from 1951 on.

[49] This sentence describes unweighted sample counts. Weighted and unweighted distributions of sample observations are usually quite similar, because the sample is designed to be "self-weighting." I am grateful to Raj Singh for a helpful discussion of the sample design of the 1984 SIPP panel. Recall that many young and elderly SIPP respondents were not employed in 1984.

employed in 1955 is likely to be strictly greater than 17%, because the SSA coverage rate in 1955 was only 82%.[50]

The internal consistency checks undertaken with the SSA administrative data consist of comparisons of the earnings and employment series on an annual basis. Overall the earnings and employment series appear to be quite consistent. In no single year does the share of observations with possible inconsistencies exceed 2%, and some of these discrepancies may be due to the differential effects of the screens used to eliminate very short employment spells.[51]

Estimated distributions of SSA-based measures of total work experience after screening out isolated employment spells of less than 2 quarters are shown in Table 5. Comparison with Table 1 reveals that the SSA distributions are shifted left on the experience axis, showing a greater number of persons with lower levels of work experience and a smaller number of persons with higher levels of work experience than the distribution obtained with the reported work experience variable in the SIPP. Despite the fact that 8.3% of SIPP records are unmatched by SSA records for the population as a whole, the SIPP work experience variable takes values between 1 and 40 years for approximately 128,694 thousand persons, while for the SSA variable the corresponding estimate is 140,741 thousand persons. In contrast, the SIPP experience variable takes values higher than 40 among an estimated 9,334 thousand persons, while for the SSA variable the corresponding estimate is 3,783 thousand persons.[52] Discrepancies between the SIPP and SSA

---

[50] SSA coverage rates are provided in footnote 54 below.

[51] These comparisons are made after the data are screened to eliminate very short spells of employment, as described above: small earnings values are set to zero, and salaried employment of one quarter is set to zero. Some of the discrepancies between SSA earnings and employment variables may be due to the differential effects of the screens. The timing of observed discrepancies suggests that employment may be recorded when an employer files an employee's social security number with the SSA, while earnings may be recorded only after a quarter of covered employment has been logged.

[52] Notice that the SIPP estimate of the total number of persons with some work experience is 138,028 thousand persons, which is significantly lower than the corresponding SSA estimate of 144,524 thousand

work experience distributions are less pronounced when limited to employed persons, but truncation of the upper tail of the SSA experience distribution remains evident.

Comparison of SIPP and SSA Work Experience Measures

The SIPP-SSA matched files allow for direct comparisons between survey data and administrative record data with respect to the duration of total work experience. Summary data on discrepancies between SIPP reported experience and SSA counted experience variables are provided in Table 6, and graphed in Chart 5.

The relative frequency with which the SSA measure exceeds the SIPP measure is greater among all observations than it is among the currently employed. This result appears to be attributable in part to the fact that young persons, who account for a larger share of the population than they do among the employed, are more likely to have worked in industries covered by SSA during their entire working lives, because SSA coverage rates were relatively high during their entire working lifetimes. Younger persons are also more likely to have held jobs whose duration was 6 months or less during summer vacations from school, and these short-lived summer jobs would be recorded in SSA administrative records, but not in the SIPP. Therefore young persons are likely to have SSA total work experience values that are significantly greater than SIPP total work experience values.

Decomposition of these results for cohorts of the employed, reported in Table 7, shows a distinctly smaller share of observations in which the SIPP reported value exceed the SSA counted

persons. Presumably this result is due to the fact that the SSA records include a large number of short employment spells that are not incorporated in the SIPP work history questions. See the discussion directly below.

value among respondents younger than 50, and a distinctly increasing share in which the SIPP reported value exceed the SSA counted value among respondents older than 50. This shift in the sign of the discrepancy between work experience estimates corresponds to a substantial increase in coverage, and especially voluntary coverage, that was instituted in 1955.[53]

Taken together these results suggest that the incidence of measurement error in the SIPP data is higher than it is in the SSA records for observations on young persons, while for observations on older persons the incidence of measurement error is higher in the SSA records than it is in the SIPP data. However on-the-job training is likely to have been minimal for young persons aged 16-21 who were employed less than 6 months, and the duration of the current or most recent job lasting 6 months or longer may be a reasonable proxy for total work experience at these young ages. The magnitudes of measurement errors in the SSA records are likely to be greater than they are in the SIPP, especially among mature respondents, because job tenure is positively correlated with age and because SSA coverage was quite incomplete for 20 of the 50 years spanned by Social Security. Therefore, on balance, these results indicate that the SIPP measures of work experience dominate the measures obtained with administrative record data.

---

[53] The share of civilian wage and salary disbursements covered by unemployment insurance increased to 81.5 percent in 1955, compared with 76.2 percent in 1954 and the previous all-time high of 77.1 percent in 1948. The large percentage increase in 1955 was caused by an expansion of coverage to include Federal, State, and local employees except for those in the Armed Forces. See the Social Security Bulletin Annual Statistical Supplement, 1974, Table 6, p. 42.

Persons who were 21 years old in 1955, at the time of this expansion in SSA coverage, were roughly 50 years old in 1984 when the SIPP survey data were collected. Therefore the cohort aged 50-55 in 1984 was the first to be significantly affected, roughly from the beginning of their adult working life, by the 1955 expansion in SSA coverage.

Simulated Experience Profiles

The quantitative implications of these differences in sources of actual work experience data are illustrated in simulated experience profiles for married women and men, generated with alternative sets of experience equation parameter estimates.[54] These experience profiles are presented in Charts 6-17, where the mnemonic *potexp* denotes potential experience, *pexpsipp* identifies simulated values based on an experience equation whose dependent variable is total work experience reported in the 1984 SIPP, and *pexpssa* denotes simulated values based on the 1984 SSA work experience variable.[55]

Among persons with sixteen or more years of schooling, simulated profiles based on the 1984 SSA experience equation are flatter than those based on the 1984 SIPP experience equation. There is clearly a systematic relationship between the downward bias of the SSA experience estimates among older cohorts, shown in Table 7, and the flatness of the SSA experience profiles shown in Charts 6-17.

Intuitively, the experience profiles shown in Charts 6-17 are cross-section "snapshot" characterizations of the relationship between age and total actual work experience at a single point in time; they do not describe the "representative individual" over the course of her or his lifetime. Rather, these profiles show how rapidly average levels of work experience increase as

---

[54] The specification of the experience equation follows that given in BLS Bulletin 2426, cited above.

[55] Experience profiles were also generated with coefficient estimates based on the entire SIPP sample, including persons who are currently employed in industries that are dominated by non-profit corporations, and persons who are currently unemployed or out of the labor force. These sample selection criteria are more comparable to those employed by Heckman. The resulting profiles are not reported here because the object of the exercise was to benchmark against results from OPT's current methodology. See Heckman (1980), "Sample Selection Bias," cited above, especially the notes to Tables 5.1 and 5.2.

one moves along the age distribution of persons who were employed in 1984.[56] Actual SSA work experience values recorded for older persons are below their true values because SSA coverage rates were relatively low during the early years of the program, and the experience profiles estimated from these data are correspondingly flatter.

Overally, experience profiles based on the 1984 SIPP and the 1984 SIPP-SSA experience equation parameter estimates are similar, but profiles based on the 1984 SSA data appear to be slightly flatter than their SIPP counterparts in most cases.[57] Experience profiles based on less-restrictive sample selection criteria and the SIPP work history variable, not displayed here, exhibit greater curvature; this result reflects lower employment rates among younger and older respondents in the population as a whole. Discrepancies between SIPP- and SSA-based profiles are greatest among highly-educated workers. Especially among women, it seems possible that in the 1930's and 1940's highly-educated employees were employed by non-profit organizations that were not covered by SSA legislation.

Predicted work experience values for older females based on the less-restrictive sample selection criteria mentioned above, not reported here, are substantially lower than those displayed there, where the sample is restricted to persons currently employed in for-profit industries. This result reflects the relatively weak lifetime labor force attachment of many older women. Declines in

---

[56] Helpful illustrations of the distinction between cross-sectional and longitudinal profiles are provided in Figures 1, 1A, 2 and 2A, and in the text that discusses these Figures, found in Ruggles and Ruggles (1977), "Anatomy of Earnings," cited above. The cross-sectional age-earnings profiles in their Figures 1 and 1A display much greater curvature than the average longitudinal age-cohort earnings profiles in their Figures 2 and 2A.

[57] Note that discrepancies between SIPP- and SSA-based profiles are greatest among highly-educated workers. Especially among women, it seems possible that highly-educated employees were likely to have been employed by non-profit organizations that were not covered by the early SSA legislation.

predicted work experience based on the less-restrictive sample are much smaller for older men, as is consistent with their stronger labor force attachment.

Estimation of Wage Rates

Recall Chart 3, in which it is shown that average wage rates and the probability of employment will vary with shifts in the wage offer distribution. Sample selection bias may arise when the probability of employment is a function of the individual's characteristics, and wage equations are estimated without taking account of the probability that an individual with a given set of characteristics will be employed.[58] Wage equations that do not take account of the probability of employment generate predicted wage rates reasonably well in the neighborhood of sample means, but may be relatively inaccurate for observations that are far from observed mean values.[59] Accuracy across the full distribution of observed wage rates is particularly desirable for the construction of the labor composition index, because it requires predicted wage rates for cells that span a broad range of values of potential experience and schooling. The implications of the coefficients obtained for predicted wage rate values across the entire age distribution of workers, when wage equations have been estimated with work experience proxies based on different sources of work experience data, and after incorporating a selection bias correction factor, is illustrated with a number of charts of wage rate profiles.

---

58 Early research in this area includes that of Cain and Watts, Gronau, Heckman, and Lewis. Heckman's work has been especially influential. See Glen C. Cain and Harold W. Watts (1973), "Toward a Summary and Synthesis of the Evidence," in Income Maintenance and Labor Supply, Cain and Watts, Eds., New York: Academic Press, pp. 328-367; Ruben Gronau (1973), "The Intrafamily Allocation of Time: The Value of Housewives' Time," American Economic Review, 63(4), pp. 634-651; and (1974), "Wage Comparisons - A Selectivity Bias," Journal of Political Economy, 82(6), pp. 1119-1145; H. Greg Lewis (1974), "Comments on Selectivity Biases in Wage Comparisons," Journal of Political Economy, 82(6), pp. 1145-1155; and James J. Heckman (1974), "Shadow Prices, Market Wages, and Labor Supply," Econometrica, 42 (4), pp. 679-694.

59 See, for example, Smith (1980), "Introduction," cited above, and Marshall and Zarkin (1987), "Effect of Job Tenure," cited above.

Selection Bias Correction

Work with the SIPP to incorporate a selection bias correction factor into the specification of the OPT wage equation, in order to take account of the probability of employment, follows the two-stage procedure of Heckman. In the first stage an additive selection-bias correction factor, λ, is derived from the predicted values of a probit equation whose dependent variable is the probability of being unemployed or out of the labor force. The probability of not being employed is modeled as a function of the determinants of offered and reservation wage rates. In the second stage of the procedure λ is entered as an additional explanatory variable in a wage equation.[60] Data on all sample observations are brought to bear in the estimation of λ, while the wage equation is estimated only with observations on persons who were employed in 1984, for whom 1984 wage data are available.[61]

The specification of the labor supply equation estimated for this project follows one developed by Heckman for the analysis of data on married women. In Heckman's model the probability of employment is represented as a function of the number of children less than 6 years old, the value of assets, the hourly wage rate of the husband, the number of years of labor market experience of

[60] For details on the estimation procedure see James Heckman (1980), "Sample Selection Bias as a Specification Error," in Female Labor Supply: Theory and Estimation, James P. Smith, Editor, Princeton, Princeton University Press, pp. 206-248; and the description of "Procedure V" in Killingsworth's 1983 taxonomy, "Second-generation studies of static models," cited above; and the attached Appendix.

[61] In earlier empirical work it was argued that estimation of a wage equation with data on employed persons alone would yield parameter estimates that were representative for employed persons, though not for the population as a whole. However this view no longer appears to be sustained in the literature. For example, in discussing predicted values of wage rates that have been estimated with an equation that incorporates a selection bias correction factor, Killingsworth notes that a predicted value obtained by setting the coefficient for λ equal to zero "is an unbiased estimate of the wage that someone with characteristics X_i in the population can earn, on average," while a predicted value that incorporates an estimated value for λ and its coefficient "is an unbiased estimate of the wage that a worker with characteristics X_i earns." Killingsworth (1983), "Second-generation studies," cited above.

the wife, and the wife's education level.[62] The specification examined here is essentially similar, except that the ratio of income from assets to the level of income below which a family is officially in poverty is substituted for the asset variable to adjust for variations in household size, a dummy variable is entered for persons who report a disability, a spline function in age is incorporated to allow the relationship to change for observations on elderly people, the wage of the spouse is given a zero value for persons living in single-headed households, and actual work experience 5 years prior to the survey is entered in place of actual or estimated current work experience.[63]

Table 14 presents results from initial probit equations whose dependent variable is the probability that the respondent is not employed. These equations are estimated by separately by gender. All explanatory variables are clearly significant, for both men and women. Notice that the signs of the estimated coefficients of these equations are opposite for males and females in the case of the intercept, the number of children younger than 6, and the wage of the spouse. And the magnitude of the coefficient associated with the normalized asset income variable is larger for females than for males. All other estimated coefficients have the same sign, and comparable magnitudes, for both genders.

---

[62] Heckman reports probit estimates for two versions of this specification. In the first, actual work experience is treated as an exogenous variable. In the second, a predicted experience is entered as a proxy for actual work experience. Heckman reports in Table 5.1 of his paper that the following variables were used to predict total work experience: "linear and squared terms for children less than six, 1967 assets, husband's age, husband's education, husband's hourly wage, wife's education, and interactions of all linear terms." OPT's current experience equation is similar to Heckman's specification, except that variables that characterize the productivity of the spouse are omitted, potential experience is included, and the functional form is a quadratic in potential experience, rather than a quadratic in number of children.

[63] The use of lagged work experience as an explanatory variable in order to avoid simultaneous equations bias is motivated by the distinction between within-sample and pre-sample experience that is made in Heckman's 1981 paper, "Heterogeneity and State Dependence," in Studies in Labor Markets, Sherwin Rosen, Ed., Chicago: University of Chicago Press, pp. 91-139. I am indebted to Jeff Zabel for this reference. The econometric motivation for this approach is discussed, for the more general case, in Phoebus J. Dhrymes (1981), Distributed Lags: Problems of Estimation and Formulation. New York: North-Holland.

Predicted values of λ, generated with estimated coefficients of the probit equations reported in Table 14, were used to generate a selection bias correction factor that is entered in the current OPT wage equation. Columns 3 and 5 of Table 15 present the resulting wage equation parameter estimates. In both the male and female wage equations, the t-statistic reported for the coefficient of λ is significant at the 5% level, providing an initial indication that the adjustment factor should be incorporated into the specification of the wage equation.[64] For both genders estimated standard errors of wage equation residuals are slightly smaller when the selection bias correction factor is included.[65]

One plausible explanation of the improved fit obtained with the selection-bias-corrected procedure is that it takes account of systematic determinants of the likelihood that a young woman is specializing in household production. The probit equation parameter estimates reported in Table 14 show that the presence of young children and the wage rate of the spouse are both negatively related to the converse of the probability of employment among men, and therefore positively related to the probability of employment. In contrast, the presence of young children and the wage rate of the spouse are positively related to the converse of the probability of employment among women, and thus negatively related to the probability of employment.

---

[64] As noted by Heckman, this t-statistic is biased upward unless the residuals of the labor supply and wage equations are uncorrelated. The formulas to be implemented in the calculation of unbiased t-statistics are given in Appendix A of Heckman (1980), "Estimation of Labor Supply Functions," cited above.

[65] Specifically, the estimated standard deviations of residual values for the wage equation for females, with and without the adjustment, are 36.8720217 and 37.1452819 respectively. For males, the estimated standard deviations of adjusted and unadjusted equation residuals are 40.3670767 and 40.7318284 respectively. These differences may be too small to be statistically significant. Formal tests of the null hypothesis that the true coefficient for lambda is equal to zero are described in Phoebus J.Dhrymes (1986), "Limited Dependent Variables," Handbook of Econometrics, Vol. III, Z. Griliches and M. Intrilligator, Eds., Amsterdam: North-Holland.

There is a positive relationship between λ and the coefficients of the probit labor supply equation, and λ enters both wage equations with a positive coefficient. Therefore the presence of young children and the wage rate of the spouse are both negatively related to wage rates observed among men, but positively related to wage rates observed among women. This result is consistent with theoretical models in which the reservation wage of young women who specialize in household production is higher than the reservation wage of young women who do not, all else equal, and positively correlated with the wage rate of the spouse.[66] It is also consistent with models in which employers do not invest in training for women whose labor force attachment is expected to be weak.

Allowing the point at which the schooling/wage rate profile intersects the wage rate axis to vary with the characteristics that are included in the labor force participation equation, through the introduction of λ, appears to permit a slightly slower increase in wage rates with increased schooling. That is, the schooling coefficients reported in Table 15 for the equations that include λ in their specification imply slightly smaller returns to schooling than those implied by coefficients of the equation estimated without λ.[67] However, under the assumptions of the former model the coefficients in the selection-bias-corrected equation are normalized by the variance of the labor supply equation, and therefore the two sets of parameter estimates are not directly comparable.

---

[66] These models are presented in the articles by Gronau and Becker, cited above.

[67] The differences between the coefficients of the categorical schooling variables in equations that include or exclude the selection bias correction factor are almost always less than two estimated standard errors apart. Conventional F tests of the joint null that the schooling coefficients of the adjusted and unadjusted wage equations are equal to one another fail to reject the null. As noted in footnote 68, standard errors estimated with OLS may be biased downward when the selection correction is appropriate for the true model, and conventional F statistics are subject to the same limitation.

Simulated Wage Rate Profiles

Wage rate profiles based on the estimated parameters of wage equations, in which the predicted values described above serve as proxies for actual work experience, have also been generated with microdata from the SIPP. Illustrative profiles based on these wage equation parameter estimates are provided in Charts 18-29. In these charts the mnemonic *wagepot* refers to profiles based on wage equations in which potential experience is used as a proxy for actual work experience. The mnemonic *wagesip* identifies profiles estimated with coefficients from an equation in which predicted SIPP work experience in 1984 is used as a proxy for actual work experience, *wagessa* identifies the equation in which predicted SSA work experience is a proxy for actual work experience, and *wageold* identifies the equation in which predicted experience is estimated with experience equation coefficients previously estimated with 1973 CPS-SSA linked microdata.

The wage rate profiles in which the work experience proxy is based on coefficient estimates from the 1973 CPS-SSA matched files exhibit fairly dramatic differences from the wage rate profiles estimated with contemporaneous work experience data, for all cases examined. These results may be telling something of a cautionary tale with respect to the unwarranted imposition of simplifying structural restrictions and/or the assumption that the experience equation coefficients are stable over time, since these wage rate profiles are clearly too steep at younger and older age ranges. Assuming that this problem also arises in other years, as it seems likely to do, these results indicate that the current labor composition index may assign weights that are too large to hours growth rates associated with younger and older workers, and weights that are too small for hours growth rates that are associated with prime-aged workers.

The incorporation of λ into the wage equation has a quantitatively smaller, but economically interesting effect on the estimated wage rate profiles (not shown here). It raises the wage rate profiles for men very slightly, and results in a distinct downward shift of the wage rate profiles of women, particularly among highly educated women. Viewed independently from the adjusted wage profiles for men, the fact that this downward shift is observed in profiles of both married and unmarried women, with and without children, might indicate that it reflects respondents' decisions to forego earnings during the early stages of their working life in order to invest in education, since persons who complete sixteen or more years of schooling are generally enrolled in school at ages 16-22.

However, highly-educated young men are also still in school during the early stages of their working lives, and in this case incorporation of λ has a much smaller effect. Since the signs of all other coefficients in the female and male labor force participation equations are identical, except for the coefficients associated with the intercept, the number of young children and the wage rate of the spouse, and the magnitudes of the same-signed coefficients are usually comparable, the differential effects of young children and the wage rate of the spouse on the gap between the market-oriented human capital accumulated by young men and young women appear to be the forces that underly the differential effects of λ in these wage rate profiles.

While the results that pertain to the specification and estimation of the labor supply equations that are used to estimate λ lend themselves to straightforward economic interpretation, further research is indicated to evaluate their robustness and statistical significance. In particular, an alternative specification of the labor supply equation that takes more detailed account of household composition and household stability might yield a better-fitting wage equation. Some of this work has been undertaken during the second stage of the project, which focuses on the estimation of extended specifications of the labor supply and wage equations with successive overlapping panels of the SIPP that span the 10-year period 1984-1993. In view of the new and challenging structure of the SIPP, it is particularly encouraging to report that the achievement and duplication of sensible empirical results has been straightforward in both stages of this research.

Conclusion

The SIPP is the only large, nationally-representative longitudinal survey for the U.S. that collects work history data on all household members of working age or older on an annual basis. Direct comparison of SIPP work experience data with administrative record data from the SSA reveals that the effect of incomplete SSA coverage at the inception of the Social Security program in 1937 is a leftward shift of the distribution of total work experience as measured in 1984, and a truncation of the distribution of the1984 SSA work experience measure at higher values. Consequently, 1984 experience profiles estimated with SSA data are flatter than experience profiles estimated with data from the 1984 SIPP. Work experience parameter estimates based on the 1984 SSA work experience measure are, nonetheless, remarkably comparable to those obtained with the 1984 SIPP reported experience variable. It is expected that discrepancies between the SIPP and SSA work experience variables are smaller for more recent years, because SSA coverage rates have been higher throughout most current workers' lifetimes.

Under the assumption that the discrepancies between wage profiles generated with the 1973 CPS-SSA matched files and those generated with contemporaneous work experience data are sustained in other years, the results reported in this paper indicate that the labor composition index obtained with the 1973 one-time match to administrative record data may overstate the rate of growth of hours worked by older workers, and understate the rate of growth of hours worked by younger workers.[68] However the surprisingly strong performance of simulated wage rate profiles in which potential experience replaces actual work experience as an explanatory variable, while schooling variables are entered separately, may indicate that cohort effects also play an important role in the determination of observed wage rates.[69]

For all cases examined here, female wage rate profiles that incorporate a selection bias correction for the probability of not working are lower than profiles that omit this adjustment. One plausible reason for the latter result is the greater likelihood that young women, rather than young men, will specialize in household production when there are young children in the household and the wage of the male spouse is relatively high. Within the framework of a behavioral model similar to those of Heckman, these variables appear to be important determinants of differences between the acceptance wages of, and the wage offers received by, young women and young men.

In summary, these results indicate that, subject to further research into the stability of selected experience and wage equations over time, the use of micro data from the SIPP could strengthen BLS labor composition indexes that span the recent past. Additional, related research on the stability of these structural relationships over time might also may help to evaluate the likelihood that CPS-SSA and SIPP-SSA matched files could be brought to bear systematically to revise labor composition indexes prior to 1984.

---

[68] Further work with the SIPP that will help evaluate this possibility is currently ongoing.

[69] Willis notes his conviction that "...an important and promising area of future research lies in the further exploration of the general equilibrium interaction of the supply and demand for human capital which has begun with the recent studies of cohort size effects," in Robert J. Willis (1986), "Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions," Chapter 10 in the Handbook of Labor Economics, Volume I, Ashenfelter and Layard, Eds, Amsterdam: North-Holland, pp. 525-602.

Bibliography

Abraham, Katherine G. and Henry S. Farber. "Job Duration, Seniority, and Earnings." American Economic Review, 1987, pp. 278-297.

Abraham, Katherine G. and Susan N. Houseman. Job Security in America: Lessons from Germany. Washington: Brookings, 1993.

Barron, John M., Dan A. Black, and Mark A. Loewenstein. "Job Matching and On-the-Job Training." Journal of Labor Economics, Vol. 7, 1989, pp. 1-19.

Baumol, William J, Sue Anne Batey Blackman, and Edward N. Wolff. Productivity and American Leadership: The Long View. Cambridge: MIT Press, 1989.

Blinder, Alan S. and Yoram Weis. “Human Capital and Labor Supply: A Synthesis.” Journal of Political Economy, Vol. 84, 1976, pp. 449-472.

Bound, John, Charles Brown, Greg J. Duncan and Willard Rodgers. "Measurement Error in Cross-Sectional and Longitudinal Labor Market Surveys: Results From Two Validation Studies." NBER Working Paper No. 2884, 1989.

Bound, John and George Johnson. "Wages in the United States During the 1980s and Beyond." In Workers and their Wages: Changing Patterns in the United States, edited by M. H. Kosters. Washington: AEI Press, 1991.

Bound, John and Alan B. Kreuger. "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" Journal of Labor Economics, 1991, pp. 1-24.

Brown, Ann and Richard Veevers. "The Use and Evaluation of Administrative Records in the Self-Sufficiency Project." 1995 Annual Research Conference Proceedings, Washington: U.S. Department of Commerce, Bureau of the Census.

Brown, James N. and Audrey Light. "Interpreting Panel Data on Job Tenure." Journal of Labor Economics, 1992, pp. 219-257.

Cain, Glen C. and Harold W. Watts. "Toward a Summary and Synthesis of the Evidence." In Income Maintenance and Labor Supply, edited by Glen C. Cain and Harold W. Watts. New York: Academic Press, 1973.

Corcoran, Mary, Greg J. Duncan, and Michael Ponze. "Work Experience and Wage Growth of Women Workers." In Five Thousand American Families - Patterns of Economic Progress: Analyses of the First Thirteen Years of the Panel Study of Income Dynamics, edited by Greg J. Duncan and James N. Morgan. Ann Arbor: Institute for Social Research, University of Michigan, 1983.

Dempster, A. P., N.W. Laird, and D.B. Rubin. "Maximum Likelihood from Incomplete Data with the EM Algorithm." Journal of Royal Statistics, Vol. 47, 1979, pp. 153-161.

Dhrymes, Phoebus J. "Autoregressive Errors in Singular Systems of Equations." Econometric Theory, Vol. 10, 1994, pp. 254-285.

__________ . "Specification Tests in Simultaneous Equations Systems." Journal of Econometrics, Vol. 64, 1994, pp. 45-76.

__________. Distributed Lags: Problems of Estimation and Formulation. New York: North-Holland, 1981.

__________. "Limited Dependent Variables." In Handbook of Econometrics, Vol. III. edited by Z. Griliches and M. Intrilligator. Amsterdam: North-Holland, 1986.

__________. "Misspecification Analysis and Errors in Variables." In Introductory Econometrics, New York: Springer-Verlag, 1978.

__________ . Topics in Advanced Econometrics, Volume I: Statistical Foundations. New York: Springer-Verlag, 1989.

__________ . Topics in Advanced Econometrics, Volume II: Linear and Nonlinear Simultaneous Equations. New York: Springer-Verlag, 1994.

Duncan, Greg J. and Daniel H. Hill. "An Investigation of the Extent and Consequences of Measurement Error in Labor-economic Survey Data." Journal of Labor Economics, 1985, pp. 508-532.

Fair, Ray C. and Kathryn Dominguez. "Effects of the Changing U.S. Age Distribution on Macroeconomic Equations." American Economic Review, 1991, pp. 1276-1294.

Godfrey, L. G. Misspecification Tests in Econometrics: The Lagrange Multiplier Principle and Other Approaches. Cambridge: Cambridge University Press, 1988.

Goldberger, Arthur. Econometric Theory. New York: Wiley, 1964.

Goldin, Claudia. "Life-Cycle Labor Force Participation of Married Women: Historical Evidence and Implications." Journal of Labor Economics, Vol. 7, 1989, pp. 20-47.

Gronau, Ruben. "Home Production - A Survey." In Handbook of Labor Economics, Vol. I, edited by Ashenfelter and Layard. Amsterdam: North-Holland, 1986.

____________. "The Intrafamily Allocation of Time: The Value of Housewives' Time." American Economic Review, Vol. 63, 1973, pp. 634-651.

_____________. "Wage Comparisons - A Selectivity Bias." Journal of Political Economy, Vol. 82, pp. 1119-1145, 1974.

Hausman, Jerry A. "Specification Tests in Econometrics." Econometrica, 1978, pp. 1251-1271.

_________________ and William E. Taylor. "Identification in Linear Simultaneous Equations Models with Covariance Restrictions: An Instrumental Variables Interpretation." Econometrica, 1983, pp. 1527-1549.

Heckman, James J. "Estimates of a Human Capital Production Function Embedded in a Life-Cycle Model of Labor Supply." In Household Production and Consumption, edited by Nestor E. Terleckyj. Studies in Income and Wealth by the Conference for Research in Income and Wealth, Vol. 40. New York: NBER and Columbia University, 1975.

_________________. "Heterogeneity and State Dependence." In Studies in Labor Markets, edited by Sherwin Rosen. Chicago: University of Chicago Press, 1981.

_________________. "Sample Selection Bias as a Specification Error." Female Labor Supply: Theory and Estimation, edited by James P. Smith. Princeton: Princeton University Press, 1980.

_________________. "Shadow Prices, Market Wages, and Labor Supply." Econometrica, Vol. 42, 1974, pp. 679-694.

Hellerstein, Judith K. and David Neumark. "Are Earnings Profiles Steeper than Productivity Profiles? Evidence from Israeli Firm-Level Data." Journal of Human Resources, 1995, pp. 89-112.

Hill, Daniel H. "Response Error Around the Seam: Analysis of Change in a Panel with Overlapping Reference Period." In Individuals and Families in Transition: Understanding Change through Longitudinal Data. Washington, D. C.: U. S. Department of Commerce, Bureau of the Census, 1988.

Iams, Howard M. "Child Care Effects on Social Security Benefits." 1991 Annual Research Conference Proceedings. Washington: U.S. Department of Commerce, Bureau of the Census, 1991.

Jorgenson, Dale W. and Barbara Fraumeni. "Investment in Education and U. S. Economic Growth." Scandinavian Journal of Economics, Vol. 94, Supplement, 1992, pp. 51-70.

Jorgenson, Dale W., Frank M. Gollop and Barbara M. Fraumeni. Productivity and U.S. Economic Growth. Cambridge, Mass.: Harvard University Press, 1987.

Juhn, Chinjui, Kevin M. Murphy and Brooks Pierce. "Wage Inequality and the Rise in Returns to Skills." Journal of Political Economy, 1993, pp. 410-442.

Killingsworth, Mark R. "Second-generation Studies of Static Labor Supply Models: Methodology and Empirical Results." In Labor Supply, Chapter 4. Cambridge: Cambridge University Press, 1983.

Killingsworth, Mark R. and James J. Heckman. "Female Labor Supply: A Survey." In Handbook of Labor Economics, Vol. I, edited by O. Ashenfelter and R. Layard, 1986, pp. 103-204.

Kosters, Marvin H. "Wages and Demographics." Workers and Their Wages, edited by Marvin H. Kosters, Washington: AEI, 1991.

Lee, Lung-Fei. "Simultaneous Equations Models with Discrete and Censored Variables." In Structural Analysis of Discrete Data with Economic Applications, edited by C. F. Manski and D. McFadden. Cambridge, Mass.: MIT Press, 1981.

_____________ and G. S. Maddala. "The Common Structure of Tests for Selectivity Bias, Serial Correlation, Heteroskedasticity and Nonnormality in the Tobit Model." International Economic Review, Vol. 26, 1985, pp. 1-20.

Lewis, H. Greg. "Comments on Selectivity Biases in Wage Comparisons." Journal of Political Economy, Vol. 82, 1974, pp. 1145-1155.

Marshall, Robert C. and Gary A. Zarkin. "The Effect of Job Tenure on Wage Offers." Journal of Labor Economics, 1987, pp. 301-324.

Martini, Alberto and Paul Ryscavage. "The Impact of Survey and Questionnaire Design on Longitudinal Labor Force Measures." In 1991 Annual Research Conference Proceedings. Washington: U.S. Department of Commerce, Bureau of the Census, 1991.

Mathiowetz, Nancy and Greg Duncan. "Out of Work, Out of Mind: Response Errors in Retrospective Reports of Unemployment." Journal of Business and Economic Statistics, 1988, pp. 221-229.

McNeil, Jack, Enrique Lamas, and Sheldon Haber. "Job Tenure, Lifetime Work Interruptions, and Wage Differentials." SIPP Working Paper No. 8711. U. S. Department of Commerce, Bureau of the Census, 1987.

Medoff, James L. and Katherine G. Abraham. "Are Those Paid More Really More Productive? The Case of Experience." Journal of Human Resources, Vol. 16, 1981, pp. 186-216.

Mincer, Jacob. "Human Capital, Technology, and the Wage Structure." Photocopy, Columbia University, 1991.

______________. Schooling, Experience and Earnings. New York: NBER and Columbia University Press, 1974.

Mincer, Jacob and Boyan Jovanovic. "Labor Mobility and Wages." In Studies in Labor Markets, edited by S. Rosen. Chicago: University of Chicago Press, 1981.

Moeller, Linda. "New Schooling Variable in the March CPS." Memo to Larry Rosenblum, Division of Productivity Research, Office of Productivity and Technology, U.S. Department of Labor, Bureau of Labor Statistics, November 3, 1993.

______________. "Labor Composition Index Number for 1992." Memo to Files, Division of Productivity Research, Office of Productivity and Technology, U.S. Department of Labor, Bureau of Labor Statistics, January 7, 1994.

______________. "Labor Composition Index for Calendar Year 1993." Memo to Files, Division of Productivity Research, Office of Productivity and Technology, U.S. Department of Labor, Bureau of Labor Statistics, January 25, 1995.

Myrskyla, Pekka, Joe Knott and Cynthia Taeuber. "Uses of Administrative Records in Censuses and Health Research: Finland and the United States." In 1995 Annual Research Conference Proceedings. Washington: U.S. Department of Commerce, Bureau of the Census, 1995.

Norwood, Janet. “Administrative Statistics: A BLS Perspective.” Journal of Business and Economic Statistics, Vol. 3, 1985, pp. 398-400.

Penceval, John. "Labor Supply of Men: A Survey." In Handbook of Labor Economics, Vol. I, edited by O. Ashenfelter and R. Layard. Amsterdam: North-Holland, 1986.

Pindyck, Robert S. and Daniel L. Rubinfeld. Econometric Models and Economic Forecasts, Second Edition. New York: McGraw Hill, 1981.

Ruggles, Nancy and Richard Ruggles. "The Anatomy of Earnings Behavior." In The Distribution of Economic Well-Being, edited by F. Thomas Juster. Studies in Income and Wealth No. 41. New York: NBER, 1977, pp. 115-158.

Ryscavage, Paul and John E. Bregger. "New Household Survey and the CPS: A Look at the Labor Force Differences." Monthly Labor Review, September 1985, pp. 3-12.

Ryscavage, Paul and John Coder. "Longitudinal vs. Retrospective Measures of Work Experience." SIPP Working Paper No. 8903. Washington: U. S. Department of Commerce, Bureau of the Census, 1989.

Ryscavage, Paul and Angela Feldman-Harkins. "A Comparison of Gross Changes in Labor Force Status from SIPP and CPS." SIPP Working Paper No. 8816. Washington: U. S. Department of Commerce, Bureau of the Census, 1988.

Schultz, Theodore W. The Economic Value of Education. New York: Columbia University Press, 1963.

Smith, James P. "Introduction." In Female Labor Supply Theory and Estimation, edited by James P. Smith. Princeton: Princeton University Press, 1980.

Social Security Bulletin, Annual Statistical Supplement, 1974. Washington: U.S. Department of Health, Education, and Welfare, Social Security Administration, 1974.

Social Security Bulletin, Annual Statistical Supplement, 1991. Washington: U.S. Department of Health, Education, and Welfare, Social Security Administration, 1991.

Stoker, Thomas M. "Simple Tests of Distributional Effects on Macroeconomic Equations." Journal of Political Economy, 1986, pp. 763-795.

Theil, Henri. Principles of Econometrics. New York: Wiley, 1971.

Thompson, Ib. "Use of Administrative Records in the Norwegian Censuses Since 1970." 1995 Annual Research Conference Proceedings. Washington: U.S. Department of Commerce, Bureau of the Census.

Topel, Robert. "Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority." Journal of Political Economy, 1991, pp. 145-176.

U.S. Department of Labor, Bureau of Labor Statistics. BLS Handbook of Methods. BLS Bulletin 2490. Washington: U.S. Government Printing Office, 1997.

U.S. Department of Labor, Bureau of Labor Statistics. Labor Composition and Productivity Growth, 1948-90. BLS Bulletin 2426. Washington: U.S. Government Printing Office, 1993.

Willis, Robert J. "Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions." In Handbook of Labor Economics, Vol. I, edited by Ashenfelter and Layard. Amsterdam: North-Holland, 1986.

Tables

TABLE 1

SIPP Reported Years Worked
Weighted Estimates, In Thousands

	Population		Employed	
Years of Experience	Females			
1 to 5	13,598	14.4	7,402	14.8
> 5 to 10	15,535	16.5	10,955	21.9
> 10 to 20	18,125	19.2	14,221	28.4
> 20 to 30	7,650	8.1	6,135	12.3
> 30 to 40	4,379	4.6	3,343	6.7
> 40 to 50	1,939	2.1	1,543	3.1
> 50	529	0.5	413	0.8
age < 16	1,426	1.5	6	0.0
age 16-21, no wrk	3,181	3.4	6	0.0
age 16-21, w wrk	6,174	6.6	4,864	9.7
never worked	5,546	5.9	0	0.0
zero, niu	16,150	17.1	1,104	2.2
total	94,233	100.0	49,996	100.0

Years of Experience	Males			
1 to 5	7,070	8.3	6,011	9.5
> 5 to 10	9,987	11.7	9,168	14.5
> 10 to 20	18,510	21.6	17,375	27.6
> 20 to 30	11,116	13.0	10,097	16.0
> 30 to 40	9,815	11.5	8,559	13.6
> 40 to 50	5,395	6.3	4,384	7.0
> 50	1,471	1.7	1,070	1.7
age < 16	1,576	1.8	11	0.0
age 16-21, no wrk	2,592	3.0	5	0.0
age 16-21, w wrk	6,735	7.9	5,522	8.8
never worked	599	0.7	0	0.0
zero, niu	10,810	12.6	846	1.3
total	85,677	100.0	63,049	100.0

Individual entries may not sum to totals, due to rounding.

TABLE 2

SIPP Calculated Years Worked
Weighted Estimates, In Thousands

	All Observations		Employed	
Years of Experience	Females			
1 to 5	8,486	9.0	2,844	5.7
> 5 to 10	8,973	9.5	4,815	9.6
> 10 to 20	10,645	11.3	7,200	14.4
> 20 to 30	4,666	5.0	3,389	6.8
> 30 to 40	2,388	2.5	1,573	3.1
> 40 to 50	453	0.5	322	0.6
> 50	44	0.0	18	0.0
age < 16	1,426	1.5	6	0.0
age 16-21, no wrk	3,181	3.4	6	0.0
age 16-21, w wrk	6,141	6.5	4,853	9.7
age ge 65	13,494	14.3	745	1.5
continuous emp.	26,854	28.5	23,312	46.6
never worked	3,307	3.5	0	0.0
zero, niu	4,029	4.3	894	1.8
dk, other	131	0.1	2	0.0
total	94,233	100.0	49,996	100.0

Years of Experience	Males			
1 to 5	2,433	2.8	1,706	2.7
> 5 to 10	2,499	2.9	1,938	3.1
> 10 to 20	3,958	4.6	3,198	5.1
> 20 to 30	2,232	2.6	1,509	2.4
> 30 to 40	2,360	2.8	1,459	2.3
> 40 to 50	859	1.0	474	0.8
> 50	142	0.2	6	0.0
age < 16	1,576	1.8	11	0.0
age 16-21, no wrk	2,592	3.0	5	0.0
age 16-21, w wrk	6,718	7.8	5,522	8.8
age ge 65	6,489	7.6	510	0.8
continuous emp.	51,430	60.0	46,036	73.0
never worked	554	0.6	8	0.0
zero, niu	1,820	2.1	660	1.0
dk, other	7	0.0	7	0.0
total	85,677	100.0	63,049	100.0

A person who is currently employed, but who has never worked on a job for a spell of 6 months or longer, is coded as "never worked." In the preceding table a single observation on an employed male who "never worked" has been recoded to the category "zero, niu." Individual entries may not sum to totals, due to rounding.

TABLE 3

SIPP Reported minus Calculated SIPP Employment Duration
Weighted Estimates, In Thousands

	All Observations		Employed	
Discrepancy	Females			
reported 4+ lt counted	3,731	3.97	2,020	4.04
reported = counted +/- 3	33,890	36.03	18,110	36.24
reported 4+ gt counted	1,962	2.09	805	1.61
continuous employment	26,850	28.55	23,310	46.65
age ge 65	11,110	11.87	729	1.46
age gt 16, lt 21	6,536	6.95	4,859	9.72
never worked	9,758	10.37	6	0.01
na	162	0.17	131	0.26
total	94,060	100.00	49,980	100.00

Discrepancy	Males			
reported 4+ lt counted	935	1.09	558	0.89
reported = counted +/- 3	14,680	17.13	9,978	15.83
reported 4+ gt counted	553	0.65	247	0.39
continuous employment	51,430	60.04	46,040	73.03
age ge 65	6,405	7.48	505	0.80
age gt 16, lt 21	7,173	8.37	5,527	8.77
never worked	4,312	5.03	11	0.02
na	177	0.21	173	0.27
total	85,660	100.00	63,030	100.00
Individual items may not sum to totals, due to rounding.				

TABLE 4

Unmatched and Matched SIPP-SSA Records
Distributions By Cohort, Unweighted

Age	Percentage by Cohort		Cohort Share
	Unmatched	Matched	of Total
16-21	13.14	86.86	13.58
21-25	13.24	86.76	9.14
26-30	8.89	91.11	11.48
31-35	6.92	93.08	10.21
36-40	6.93	93.07	8.71
41-45	7.89	92.11	7.35
46-50	7.78	92.22	6.80
51-55	8.39	91.61	6.31
56-60	6.65	93.35	6.29
61-65	6.35	93.65	6.03
66-70	4.99	95.01	5.09
71+	3.34	96.66	9.00

TABLE 5

SSA Counted Years Worked
Weighted Estimates, In Thousands

	All Observations		Employed	
Years of Experience	Females			
1 to 5	14,910	15.8	4,889	9.8
> 5 to 10	19,071	20.2	11,598	23.2
> 10 to 20	25,309	26.9	17,680	35.4
> 20 to 30	10,382	11.0	6,593	13.2
> 30 to 40	3,402	3.6	2,052	4.1
> 40 to 50	387	0.4	262	0.5
> 50	0	0.0	0	0.0
age < 16	1,361	1.4	6	0.0
age < 16, nonmtch	64	0.1	0	0.3
age 16-21, w wrk	8,986	9.5	4,724	9.4
age 16-21, nonmtch	369	0.4	147	0.3
age > 21, nonmtch	2,308	2.4	1,301	2.6
niu	7,382	8.2	743	1.5
total	94,233	100.0	49,996	100.0

Years of Experience	Males			
1 to 5	4,626	5.0	2,677	4.2
> 5 to 10	11,531	13.5	9,596	15.2
> 10 to 20	21,538	25.1	18,519	29.4
> 20 to 30	16,214	18.9	12,477	19.8
> 30 to 40	13,758	16.1	9,440	15.0
> 40 to 50	3,396	4.0	2,129	3.4
> 50	0	0.0	0	0.0
age < 16	1,500	1.8	11	0.0
age < 16, nonmtch	76	0.1	0	0.0
age 16-21, w wrk	8,817	10.3	5,264	8.3
age 16-21, nonmtch	509	0.6	263	0.4
age > 21, nonmtch	2,682	3.1	2,156	3.4
niu	1,389	1.6	516	0.8
total	85,677	100.0	63,049	100.0

These calculations are based on SSA administrative record data, and include only employment spells of 2 quarters or longer. Individual entries may not sum to totals, due to rounding.

TABLE 6

SIPP and SSA Work Experience Measures
Weighted Discrepancies Between Estimates, In Thousands

	All Observations		Employed	
Discrepancy	Females			
SSA 4+ > SIPP	22,425	23.8	10,859	21.7
Difference < 4	37,982	40.3	22,774	45.6
SIPP 4+ > SSA	15,367	16.3	10,185	20.4
Non-match	2,741	2.9	1,448	2.9
SIPP Vbl Na	15,718	16.7	4,730	9.5
Total	94,233	100.0	49,996	100.0

Discrepancy	Males			
SSA 4+ > SIPP	25,303	29.5	15,498	24.6
Difference < 4	31,881	37.2	28,304	44.9
SIPP 4+ > SSA	14,333	16.7	11,552	18.3
Non-match	3,267	3.8	2,419	3.8
SIPP Vbl Na	10,892	12.7	5,275	8.4
Total	85,677	100.0	63,049	100.0

The category "SIPP Vbl Na" includes all observations on persons younger than 21, for whom work history data are either unavailable, or potentially incomplete, and all observations on persons who have never worked. For all observations, these two categories account for 6,431 observations, of which 728 were not matched to SSA records. Among the employed these categories account for 2,409 observations, including 333 observations that were not matched to SSA records. Individual entries may not sum to totals, due to rounding.

TABLE 7

SIPP and SSA Work Experience Estimates, Employed Persons
Discrepancies by Cohort, Weighted Sample Observations

Age	Percentage by Cohort				Cohort Share of Sample
	SIPP < SSA	SIPP = SSA	SIPP > SSA	No Match	
	Females				
16-21	0.00	0.00	0.00	3.00	9.75
21-25	26.67	66.31	3.50	3.52	15.13
26-30	28.55	61.70	7.37	2.38	14.80
31-35	28.37	55.86	13.28	2.49	13.09
36-40	24.78	52.90	19.57	2.75	11.07
41-45	23.76	48.11	24.62	3.52	9.63
46-50	26.02	35.68	35.22	3.07	7.40
51-55	18.70	35.86	42.42	3.01	6.80
56-60	15.43	36.69	45.37	2.51	6.10
61-65	11.02	26.00	60.01	2.96	3.80
66-60	6.56	13.79	77.47	2.17	1.43
71+	6.49	8.07	83.46	1.98	0.99

Age	Males				
16-21	0.00	0.00	0.00	4.75	8.78
21-25	33.06	58.42	3.61	4.91	13.68
26-30	33.68	56.28	6.57	3.47	15.14
31-35	33.50	52.34	11.45	2.71	13.18
36-40	27.96	51.66	16.80	3.57	11.77
41-45	29.87	49.68	16.14	4.40	9.15
46-50	28.35	45.84	21.41	4.39	7.88
51-55	13.00	47.63	35.97	3.41	7.12
56-60	13.71	35.05	47.58	3.66	6.46
61-65	9.22	29.79	57.59	3.41	4.14
66-60	6.92	18.38	69.94	4.76	1.45
71+	0.00	7.07	92.30	0.63	1.25

Unweighted percentages of SIPP records that can be matched to SSA administrative records for the age group 16-21 are 87.52% among women, and 85.03% among men. Information on the duration of the current job, or last job held, is available, but the SIPP total work experience variable is not consistently available for persons in this age group. Therefore no comparison is made in this Table. Shares may not sum to 100%, due to rounding.

TABLE 8

Distributions of Alternative Work Experience Measures, In Thousands
Weighted Observations used to Estimate Experience Equation

	Proxy		SIPP		SSA	
Experience	Females					
0 to 5	2,235	36.1	1,393	24.2	1,131	20.2
6 to 10	1,493	24.1	1,514	26.3	1,501	26.8
11 to 15	942	15.1	997	17.4	1,164	20.8
16 to 20	1,247	20.1	660	11.5	802	14.3
21 to 25	228	3.7	433	7.5	418	7.5
26 to 30	32	0.5	248	4.3	242	4.3
31 to 35	14	0.2	208	3.6	192	3.4
36 to 40			139	2.4	80	1.4
41 to 45			93	1.6	35	0.6
46 to 50			38	0.7	12	0.2
> 50			23	0.3	14	0.3
total	6,191		5,746		5,591	
missing			445		600	
Experience	Males					
0 to 5	2,481	31.4	1,255	17.1	1,051	14.9
6 to 10	1,378	17.4	1,223	16.7	1,225	17.4
11 to 15	1,066	13.5	1,185	16.1	1,141	16.2
16 to 20	849	10.7	933	12.7	980	13.9
21 to 25	794	10.1	637	8.7	690	9.8
26 to 30	785	9.9	554	7.5	714	10.1
31 to 35	469	5.9	563	7.7	622	8.8
36 to 40	70	0.9	463	6.3	256	3.6
41 to 45	6	0.1	335	4.5	206	2.9
46 to 50			119	1.6	78	1.1
> 50			77	1.0	86	1.2
total	7,898		7,344		7,049	
missing			554		849	
Individual entries may not sum to totals, due to rounding.						

Tables 9A-9C: Mnemonics for Equations Estimated

Table 9A

Mnemonic	Experience Equation, Variable Definitions
SIPPYR	dependent variable, years of total work experience reported in the SIPP
SSAYR	dependent variable, number of years with 2 or more quarters employment in SSA data
SSAQTR	dependent variable, number of quarters of employment recorded in SSA data
potexp	potential experience: age-years of schooling completed-6.
potexpsq	squared potential experience
married	1 if ever married, 0 otherwise
pomar	potexp*married
posqmar	pomar squared
potkid1	1 if 1 child present in family, 0 otherwise
potkid23	1 if 2 or 3 children present in family, 0 otherwise
potkid4	1 if 4 or more children present in family, 0 otherwise
s0to4	1 if less than 4 years of schooling completed, 0 otherwise
s5to8	1 if 5 to 8 years of schooling completed, 0 otherwise
s12	1 if 12 years of schooling completed, 0 otherwise
s13to15	1 if 13 to 15 years of schooling completed, 0 otherwise
s16	1 if 16 years of schooling completed, 0 otherwise
s17up	1 if 17 or more years of schooling completed, 0 otherwise
pos0to4	s0to4*potexp
pos5to8	s5to8*potexp
pos12	s12*potexp
pos13to15	s13to15*potexp
pos16	s16*potexp
pos17up	s17up*potexp
potexpnp	potexp if currently employed in an industry dominated by nonprofits, 0 otherwise
prev	1 if year of first job was prior to 1937
black	1 if respondent is Black, 0 otherwise
hisp	1 if respondent is Hispanic, 0 otherwise

Table 9B

Mnemonic	Wage Rate Equation, Variable Definitions
whhr	dependent variable, annual earnings divided by annual hours worked
SIPPYR	indicates experience proxy estimated with SIPP data on years worked
SSAYR	indicates experience proxy estimated with SSA data on years worked 2+ quarters
SSAQTR	indicates experience proxy estimated with SSA data on total quarters worked
OLDSSA	indicates experience proxy estimated with data from 1973 CPS-SSA matched file
POTEXP	indicates experienc proxy is estimated as age-years of school completed-6
s0to4	1 if less than 4 years of schooling completed, 0 otherwise
s9to11	1 if 5 to 8 years of schooling completed, 0 otherwise
s12	1 if 12 years of schooling completed, 0 otherwise
s13to15	1 if 13 to 15 years of schooling completed, 0 otherwise
s16	1 if 16 years of schooling completed, 0 otherwise
s17up	1 if 17 or more years of schooling completed, 0 otherwise
experience	proxy for actual work experience
expsq	squared proxy for actual work experience
black	1 if respondent is Black, 0 otherwise
hisp	1 if respondent is Hispanic, 0 otherwise
ptime4	1 if usual hours worked when employed is less than 35, 0 otherwise
vet	1 if served in armed forces, 0 otherwise
city	1 if a resident of a central city within an SMSA
neweng	1 if a resident of the New England region, 0 otherwise
midatl	1 if a resident of the Mid Atlantic region, 0 otherwise
encent	1 if a resident of the East North Central region, 0 otherwise
wncent	1 if a resident of the West North Central region, 0 otherwise
satl	1 if a resident of the South Atlantic region, 0 otherwise
escent	1 if a resident of the East South Central region, 0 otherwise
wscent	1 if a resident of the West South Central region, 0 otherwise
mountain	1 if a resident of the Mountain region, 0 otherwise
λ	selection bias correction factor

Table 9C

	Labor Supply Equation, Variable Definitions
working	dependent variable, 0 if wage rate can be calculated for observation, 1 otherwise**
young	number of children 6 years old or younger present in family
assypov	income from financial assets during 1984, divided by family's poverty cutoff level
othwage	wage rate of spouse if present, 0 otherwise
last5	years of work experience accumulated 5 years prior to interview
retire	years of work experience - 20 if difference is positive, 0 otherwise
schoolyr	years of schooling completed
splold	age - 65 if difference is positive, 0 otherwise
notwell	1 if a physical, mental or other health condition limits work, 0 otherwise
** SAS estimates the probability that the event with the lowest coded value occurs	

TABLE 9

1984 Experience Equation Parameter Estimates
Specific Source of Experience Data Shown in Column Headings

	Women			Men		
variable	SIPPYR	SSAYR	SSAQTR	SIPPYR	SSAYR	SSAQTR
	R^2=.58 F=301.90	R^2=.52 F=231.24	R^2=0.61 F=363.73	R^2=.88 F=2763.6	R^2=.74 F=1022.1	R^2=.82 F=1918.5
intercept	-1.151**	0.832*	-2.8509*	-1.274**	0.553	-1.4073
potexp	0.977**	0.949**	3.6722**	1.071**	1.124**	4.1425**
potexpsq	-0.002	-0.004**	-0.0126**	-0.004**	-0.007**	-0.0209**
married	0.47700	0.703*	2.7757**			
pomar	-0.050	-0.058	-0.4507**			
posqmar	-0.005**	-0.004**	-0.0006			
potkid1	-0.089**	-0.060**	-0.1771**			
potkid23	-0.143**	-0.111**	-0.4639**			
potkid4	-0.198**	-0.201**	-0.8435**			
s0to4	-5.200*	-2.240	-8.0759	-1.229	-1.976	-1.8783
s5to8	-4.072**	-2.157*	-6.7611	-2.187**	-3.789**	-12.5838**
s12	1.783**	1.864**	9.0017**	1.309**	2.238**	6.9897**
s13to15	3.151**	4.087**	16.7200**	2.533**	3.523**	11.2252**
s16	3.884**	5.432**	21.2730**	3.142**	5.048**	15.8537**
s17up	4.292**	6.074**	25.5769**	4.014**	7.302**	20.6890**
pos0to4	0.139	-0.047	-0.1723	-0.096**	-0.144**	-0.6162**
pos5to8	0.104**	0.025	-0.0965	0.020**	0.024	0.0641
pos12	-0.011	-0.007	-0.1505**	0.011	-0.055**	-0.0767
pos13to15	-0.040	-0.108**	-0.4464**	0.026*	-0.057**	-0.0468
pos16	-0.067**	-0.193**	-0.8095**	0.011	-0.064**	-0.1002
pos17up	-0.011	-0.236**	-1.0269**	-0.018	-0.280**	-0.5227**
potexpnp	-0.032**	-0.073**	-0.2560**	0.005	-0.134**	-0.4637**
prev	0.995**	0.179*	-0.2639	1.519**	-0.107*	-0.5313**
black	-0.084	-0.260	-0.9638	-1.874**	-1.700**	-4.4481**
hisp	-0.834**	-0.962**	-4.5128**	-1.162**	-2.422**	-9.8445**

** denotes significance at the 5% level,* denotes significance at the 10% level.
Mnemonics are defined in Appendix A. Specification and sample selection criteria are comparable to current OPT procedures.

TABLE 10

1984 Wage Equation Parameter Estimates
Specific Source of Experience Data Shown in Column Headings

	Women				
variable	SIPPYR	SSAYR	SSAQTR	OLDSSA	POTEXP
	R^2=0.26 F=88.83	R^2=0.27 F=89.47	R^2=0.24 F=93.87	R^2=0.26 F=86.80	R^2=0.27 F=92.40
intercept	1.2564**	1.1024**	1.1386**	1.2561**	1.3121**
s0to4	-0.1353**	-0.1100	0.0543	-0.1155*	-0.1434**
s9to11	0.1120**	0.1180**	0.1130**	0.1081**	0.0715**
s12	0.2341**	0.2325**	0.2423**	0.2398**	0.2039**
s13to15	0.4056**	0.3964**	0.3914**	0.4352**	0.3950**
s16	0.5681**	0.5538**	0.5528**	0.5880**	0.5709**
s17up	0.6485**	0.6414**	0.6424**	0.6810**	0.6555**
exp	0.0362**	0.0521**	0.0135**	0.0278**	0.0278**
expsq	-0.0008**	-0.0012**	-0.000090**	-0.0000010**	-0.0005**
black	-0.0783**	-0.0759**	-0.0776**	-0.0765**	-0.0889**
hisp	-0.0090	-0.0041	-0.0121	-0.0178	-0.0360**
ptime4	-0.2401**	-0.2368**	-0.1827**	-0.2419**	-0.2400**
city	0.0410**	0.0414**	0.0473**	0.0395**	0.0473**
neweng	0.0325	0.0313*	0.0532**	0.0337	0.0358
midatl	0.0318*	0.0318*	0.0461**	0.0319*	0.0366**
encent	-0.0363**	-0.0356**	-0.0207	-0.0361**	-0.0342**
wncent	-0.0654**	-0.0638**	-0.0434*	-0.0637**	-0.0585**
satl	-0.0687**	-0.0688**	-0.0460**	-0.0671**	-0.0631**
escent	-0.1330**	-0.1316**	-0.1282**	-0.1293**	-0.1292**
wscent	-0.0476**	-0.0495**	-0.0281	-0.0462	-0.0469**
mountain	-0.0513	-0.0491	-0.05980*	-0.0508	-0.0454

** denotes significance at the 5% level,* denotes significance at the 10% level.
Mnemonics are defined in Appendix A. Specification and sample selection criteria are comparable to current OPT procedures. Selection bias corrected results are in Table 15.

TABLE 10
Continued

1984 Wage Equation Parameter Estimates
Specific Source of Experience Data Shown in Column Headings

Men					
variable	SIPPYR	SSAYR	SSAQTR	OLDSSA	POTEXP
	R^2=0.40 F=209.13	R^2=0.40 F=208.69	R^2=0.39 F=231.56	R^2=0.40 F=213.35	R^2=0.40 F=215.31
intercept	1.3108**	1.0821**	1.1687**	1.2514**	1.2973**
s0to4	-0.0682	-0.0776	-0.1387**	-0.0713	-0.0717
s9to11	0.1917**	0.1906**	0.1501**	0.1775**	0.1651**
s12	0.3127**	0.2878**	0.2904**	0.2977**	0.3104**
s13to15	0.4220**	0.3938**	0.3824**	0.4299**	0.4508**
s16	0.6843**	0.6289**	0.6446**	0.6514**	0.7222**
s17up	0.7524**	0.7030**	0.7321**	0.7370**	0.8000**
exp	0.0477**	0.0698**	0.0183**	0.0446**	0.0517**
expsq	-0.0007**	-0.0012**	-0.00009**	-0.0000006**	-0.0008**
black	-0.2047**	-0.2075**	-0.1651**	-0.2072**	-0.2031**
hisp	-0.0610**	-0.0586**	-0.0073**	-0.0616**	-0.0792**
ptime4	-0.3277**	-0.3109**	-0.1540**	-0.3164**	-0.3140**
vet	0.0065	0.0249*	0.0251**	0.0127	0.0073
city	-0.0110	-0.0132	-0.0204*	-0.0094	-0.0076
neweng	0.0162	0.0131	0.0016	0.0191	0.0166
midatl	0.0140	0.0173	-0.0069	0.0184	0.0216
encent	-0.0218	-0.0171	-0.0373**	-0.0196	-0.0177
wncent	-0.0641**	-0.0639**	-0.0970**	-0.0610**	-0.0574**
satl	-0.1065**	-0.1076**	-0.1198**	-0.1112**	-0.1088**
escent	-0.1340**	-0.1348**	-0.1648**	-0.1325**	-0.1344**
wscent	-0.0046	-0.0079	-0.0031	-0.0052	-0.0063
mountain	-0.0043	-0.0071	-0.0556*	-0.0068	-0.0064
prev			-0.0120**		

** denotes significance at the 5% level, * denotes significance at the 10% level. Mnemonics are defined in Appendix A. Specification and sample selection criteria are comparable to current OPT procedures. Selection bias corrected results are in Table 15.

TABLE 11

F Statistics for the Null that 1984 Experience Equation Coefficients
Are Equal to those Estimated for 1973
Specific Source of Experience Data Shown in Column Headings

	Women			Men		
variable	SIPPYR	SSAYR	SSAQTR	SIPPYR	SSAYR	SSAQTR
intercept	9.57**	0.34	8.47**	48.61**	0.07	15.07**
potexp	2.18	4.92**	11.20**	30.94**	40.63**	15.27**
potexpsq	24.85**	13.62**	147.78**	22.80**	9.58**	1011.52**
married	0.00	0.24	0.30			
pomar	9.78**	11.47**	56.23**			
posqmar	35.03**	35.56**	15.31**			
potkid1	1.43	1.90	40.01**			
potkid23	23.95**	68.99**	353.70**			
potkid4	11.91**	14.24**	292.51**			
s0to4	0.04	0.70	9.12**	8.50**	3.02*	14.60**
s5to8	0.85	0.25	15.12**	2.71*	0.56	100.20**
s12	1.97	3.05*	8.60**	0.14	4.31**	2.79*
s13to15	2.10	11.74**	18.38**	0.02	5.02**	0.99
s16	2.23	18.04**	22.13**	2.33	7.70**	0.95
s17up	0.05	6.83**	12.41**	0.02	30.70**	7.04**
pos0to4	0.96	1.29	1.87	4.80**	7.38**	18.18**
pos5to8	2.89*	0.09	3.25*	0.35	0.08	1.03
pos12	0.03	0.15	10.23**	2.54	7.48**	0.67
pos13to15	7.93**	0.05	128.34**	8.48**	4.23**	2.50
pos16	6.36**	1.88	189.68**	9.11**	0.74	12.97**
pos17up	28.99**	0.57	239.66**	21.17**	30.93**	122.03**
potexpnp	9.00**	0.13	52.48**	47.22**	8.77**	84.40**
prev	79.88**	3.29*	0.75	1063.39**	11.98**	10.13**
black	0.07	0.82	0.99	86.48**	40.94**	27.90**
hisp	3.88**	6.57**	11.44**	27.15**	67.89**	103.85**
all	27.10**	36.00**	8377.63**	270.70**	48.53**	4101.14**

** denotes rejection at the 5% significance level or lower,
* denotes rejection at the 10% significance level or lower.
Mnemonics are defined in Appendix A. Specification and sample selection criteria are comparable to current OPT procedures. Selection bias corrected coefficients are reported in Table 15.

TABLE 12

F Statistics for the Null that 1984 Wage Equation Coefficients
Estimated with 1984 SIPP Data Are Equal to those Estimated
with 1973 SIPP-SSA Proxy and 1984 CPS Data
Specific Source of Experience Data Shown in Column Headings

	Females			
	SIPPYR	SSAYR	SSAQTR	1973 Proxy
intercept	29.32**	1.17	4.70**	28.35**
s0to4	1.68	2.76*	17.10**	2.48
s9to11	0.66	1.00	0.77	0.49
s12	0.18	0.22	0.02	0.05
s13to15	0.18	0.02	0.00	1.84
s16	0.80	0.23	0.22	2.15
s17up	2.23	2.82*	1.43	0.43
exp	91.89**	1.66	3150.57**	308.76**
expsq	145.78**	9.86**	53993.4**	242185843**
ptime4	19.52**	17.30**	0.00	20.70**
city	31.09**	30.75**	29.89**	32.12**
neweng	3.36*	3.21*	7.88**	3.50**
midatl	18.87**	18.92**	29.60**	18.81**
encent	0.35	0.31	0.07	0.33
wncent	0.02	0.04	1.05	0.04
satl	3.28*	3.27*	9.95**	3.56**
escent	1.26	1.15	1.09	0.98
wscent	2.78*	2.51	7.14**	2.96**
mountain	0.02	0.01	0.15	0.02
vet				
all	23.74**	10.61**	36140.6**	36860267**

* denotes rejection at the 10% significance level or lower,
** denotes rejection at the 5% significance level or lower.
Mnemonics are defined in Appendix A. Specification and sample selection criteria are comparable to current OPT procedures. Selection bias corrected coefficients are reported in Table 15.

TABLE 12
Continued

F Statistics for the Null that 1984 Wage Equation Coefficients Estimated with 1984 SIPP Data Are Equal to those Estimated with 1973 SIPP-SSA Proxy and 1984 CPS Data
Specific Source of Experience Data Shown in Column Headings

	Males			
	SIPPYR	SSAYR	SSAQTR	1973 Proxy
intercept	63.82**	0.43	12.63**	35.98**
s0to4	2.13	1.64	0.02	1.98
s9to11	6.53**	6.39**	1.30	4.23**
s12	3.45**	0.75	1.11	1.61
s13to15	7.94**	3.11*	2.09	9.79**
s16	20.65**	6.93**	12.24**	11.75**
s17up	0.29	1.06	0.01	0.00
exp	545.05**	25.26**	14451.2**	849.32**
expsq	1054.6**	63.91**	2300736**	3.40947E9**
ptime4	52.78**	38.48**	11.91**	43.70**
city	31.67**	33.64**	48.22**	30.52**
neweng	6.68**	6.09**	5.18**	7.31**
midatl	14.13**	15.46**	8.46**	16.05**
encent	0.17	0.02	1.96	0.08
wncent	0.21	0.22	0.82	0.35
satl	0.06	0.10	1.05	0.25
escent	1.26	1.33	5.93	1.15
wscent	1.96	1.56	2.49	1.90
mountain	0.84	0.71	0.24	0.73
vet	5.22**	0.93	1.12	3.41*
all	92.48**	16.74**	170328.7**	929368400**

* denotes rejection at the 10% significance level or lower
** denotes rejection at the 5% significance level or lower
Mnemonics are defined in Appendix A. Specification and sample selection criteria are comparable to current OPT procedures. Selection bias corrected coefficients are reported in Table 15.

TABLE 13

F Statistics for the Null that 1984 Experience and Wage Equation Coefficients
Estimated with 1984 SIPP and1984 SSA Data are Equal

Experience Equation			Wage Equation		
	Females	Males		Females	Males
intercept	4.34**	11.72**	intercept	10.75**	17.31**
potexp	0.12	3.53*	s0to4	0.09	0.14
potexpsq	0.74	38.35**	s9to11	0.52	0.03
married	0.02		s12	0.22	0.44
pomar	0.01		s13to15	0.04	0.54
posqmar	0.02		s16	0.01	1.78
potkid1	0.72		s17up	0.01	1.30
potkid23	3.43**		exp	19.50**	45.81**
potkid4	0.32		expsq	15.63**	45.37**
s0to4	0.03	0.15	ptime4	0.02	0.45
s5to8	3.02**	2.99*	city	0.00	0.01
s12	0.11	0.83	neweng	0.00	0.01
s13to15	1.01	1.04	midatl	0.01	0.01
s16	2.19	6.03**	encent	0.00	0.02
s17up	3.20**	16.12**	wncent	0.01	0.00
pos0to4	0.43	2.50	satl	0.00	0.01
pos5to8	5.00**	0.54	escent	0.00	0.00
pos12	0.05	2.97*	wscent	0.00	0.01
pos13to15	2.15**	5.31*	mountain	0.01	0.01
pos16	7.02**	4.14**	vet		1.08
pos17up	16.48**	44.69**	all	2.06**	5.00**
potexpnp	8.14**	19.74**			
prev	38.66**	495.18**			
black	0.31	1.21			
hisp	0.05	14.74**			
all	13.55	7.63**			

** denotes significance at the 5% confidence level,
* denotes significance at the 10% confidence level.
Mnemonics are defined in Appendix A. Specification and sample selection criteria are comparable to current OPT procedures. Selection bias corrected coefficients are reported in Table 15.

TABLE 14

Preliminary Probit Equation Estimates of Probability of Not Working
For Construction of Heckman-Style Correction Factor for Sample Truncation

Variable	Women	Men
log likelihood	-47,357,733	-38,305,933
intercept	0.0934**	-0.5293**
young	0.1571**	-0.1620**
assypov	0.2253**	0.0854**
othwage	0.0162**	-0.0117**
last5	-0.0531**	-0.0235**
retire	0.0610**	0.0311**
schoolyr	-0.0375**	-0.0145**
splold	0.1885**	0.1815**
notwell	0.9404**	0.8848**

**denotes significance at the 5% level or lower.
Mnemonics are defined in Appendix A.

TABLE 15

1984 SIPP Wage Equation with Selection Bias Correction

	Women		Men	
Variable	No Adj.	With Adj.	No Adj.	With Adj.
	R^2=0.27 F=164.27	R^2=0.28 F=165.05	R^2=0.31 F=206.06	R^2=0.32 F=209.14
intercept	1.2523**	1.0676**	1.4368**	1.0758**
s0to4	0.0135	0.0234	-0.0486	-0.0205
s9to11	0.1419**	0.1142**	0.1833**	0.1425**
s12	0.2840**	0.2424**	0.3298**	0.2782**
s13to15	0.4185**	0.3217**	0.4279**	0.3456**
s16	0.5783**	0.4716**	0.6601**	0.5670**
s17up	0.6725**	0.5566**	0.7056**	0.6003**
experience	0.0332**	0.0315**	0.0307**	0.0275**
expsq	-0.0009**	-0.0010**	-0.0004**	-0.0003**
black	-0.0701**	-0.0713**	-0.1097**	-0.1102**
hisp	-0.0168	-0.0191	0.0065	-0.0122
ptime4	-0.3471**	-0.3184**	-0.4085**	-0.3637**
vet			0.0371**	0.0395**
city	0.0528**	0.0530**	-0.0093	-0.0054
neweng	0.0642**	0.0537**	-0.0013	0.0044
midatl	0.0644**	0.0591**	0.0175	0.0143
encent	-0.0261*	-0.0302*	-0.0305*	-0.0310*
wncent	-0.0437**	-0.0468**	-0.0478**	-0.0496**
satl	-0.0456**	-0.0478**	-0.0743**	-0.0707**
escent	-0.1283**	-0.1288**	-0.1266**	-0.1312**
wscent	-0.0388*	-0.0343*	-0.0255	-0.0242
mountain	-0.0308	-0.0303	-0.0434	-0.0360
λ		0.2233**		0.2964**

** denotes significance at the 5% confidence level,
* denotes significance at the 10% confidence level.
Mnemonics are defined in Appendix A.

The coefficients reported here differ somewhat from those reported in Table 10. For the estimates reported in Table 10, actual experience is predicted with a sample that is restricted to persons who were employed during the interview in which work history data were collected, and this same sample selection criterion was applied to the sample with which the wage equation is estimated. Persons employed in industries that were not covered by Social Security legislation are omitted from the estimating sample with which coefficients in Table 10 were estimated, for comparability with OPT's current procedures. For the coefficient estimates reported in Table 14, the sample with which actual experience is predicted includes all adults, the wage equation selection criterion is that the person worked at least one hour during calendar year 1984, and persons employed in industries dominated by non-profit corporations and domestic services are not deleted from the estimating sample. The latter sample selection criteria are consistent with the sample selection criteria used to estimate the total number of hours worked during calendar year 1984.

Appendix A: Correction for Selection Bias in Wage Equation

A lucid explanation of the problem of selection bias is provided by Penceval; his discussion may be summarized as follows.[70] Consider a sub-population of N persons whose observed values for all explanatory variables entered in a wage equation, represented by the matrix X, are identical. Let the reservation wage be a linear function of X and a stochastic term, denoted e. Then the true reservation wage rates for this sub-population is:

$$w^* = a + X\,b + e.$$

Here w^* is (Nx1), X is (NxG), and e is a (Nx1) random vector whose elements are assumed to be distributed normally, with mean zero. Notice that e is the sole source of differences among the reservation wages of the persons in this sub-population by assumption, and that the shape of the distribution of the elements of w* around the mean reservation wage for this group is identical to the shape of the distribution of e.

In general, all of the persons in this sub-population are not employed. Let the subscript i denotes an individual observation. Assuming a zero unemployment rate for simplicity, those who do not participate in the labor force are persons whose reservation wage w_i^* is above the market wage rate, denoted w, that is available to persons with characteristics X. Then if all observations in this group are from persons with the same observable characteristics, so that $X_i = X$ for all i, there is some value e^* such that w is greater than w_i^* if and only if e_i is less than e^*. That is, person i is employed if and only if e_i is less than e^*.

In this context OLS procedures violate standard assumptions under which OLS yields best linear unbiased coefficient estimates. Since e_i has a mean value of zero in the population as a whole, the mean value of e_i for persons who are employed is less than zero. It is not possible to remedy the situation by simply shifting the intercept, since the residuals for the employed are not distributed symmetrically around their mean value, as shown in Chart 3.

Furthermore, the vector X is not orthogonal to the residuals associated with observations on the employed. To see why the orthogonality assumption is violated, assume that n of the N persons in the population are employed.[71] Let e_e be the sub-vector of e that corresponds to these n persons, and let X_e be the corresponding submatrix of X. Each (nx1)-element column vector x_e of X_e consists of observations on a single explanatory variable. For each column x_e in X_e:

$$\begin{aligned} E(x_e'e_e) &= E(x_e'e_1 + \dots + x_e'e_n) \\ &= Ex_e'(e_1 + \dots + e_n) \\ &= x_e'ne_m, \end{aligned}$$

where $e_m=(1/n)(e_1 + \dots + e_n)$. The mean value of the elements of e among the employed, e_m, is strictly less than zero. In general, therefore, $E(X_e'e_e)$ is unequal to zero.

These violations of the standard assumptions under which OLS coefficient estimates are best linear unbiased estimates mean that wage equation parameter estimates based exclusively on data from the

[70] John Penceval (1986), “Labor Supply of Men,” cited above.

[71] Recall that orthogonality means E(X'e)=0.

employed will generally be biased and inconsistent. Of the two properties inconsistency is the most important, because it means that the estimated parameters will not converge to the true parameters no matter how large the sample size.

One approach to this problem that has been adopted in the literature involves modeling the labor supply decision explicitly as a separate equation with its own stochastic process. In this formulation an equation for the reservation wage and a labor supply equation may be determined simultaneously, or they may form a recursive system, depending on whether or not the stochastic processes associated with the two equations are correlated, and on whether or not both endogenous variables appear on the right-hand-sides of both equations.

In the case in which the true model is recursive, which is to say that the two stochastic processes are uncorrelated and one of the two endogenous variables does not appear on the right-hand-side of either equation, estimation of the wage equation through OLS procedures yields unbiased and consistent, although inefficient, parameter estimates. But in the case in which the true system is determined simultaneously OLS procedures result in parameter estimates that are biased and inconsistent. Maximum likelihood procedures are indicated in the second case.[72]

Although the exact estimation strategies differ in accordance with the model specified, the following prototype, which follows the work of Heckman, is frequently adopted when the hypothesis of a recursive structure is to be tested. To review Heckman's approach, retain the simplifying assumption that the individuals within the population examined are identical with respect to their observable characteristics.[73] Let the wage rate received by employed persons be a function of the variables Z, and let the wage offer be modeled as:

$$w_i = a' + Z\, d + u_i.$$

Assume again that the distribution of w_i is identical to the distribution of u_i by virtue of the simplifying assumption that Z_i=Z for all i.

Assume that X and Z have some elements in common, and let the matrix of explanatory variables that enter both equations be denoted C. Then X = (S,C) Z=(D,C), and the wage offer and reservation wage equations respectively may be written:

$$w_i = a' + D\, d_d + C\, d_c + u_i,$$
$$w_i^* = a + S\, b_s + C\, b_c + e_i.$$

In this revised context person i will be employed and w_i will be observed if and only if:

$$w_i - w_i^* = a' - a + D\, d_d - S\, b_s + C\,(d_c - b_c) + v_i \geq 0,$$

where $v_i = u_i - e_i$. Notice that the explanatory variables in the equation for w_i are a subset of the explanatory variables in the equation for $w_i - w^*_i$. If the correlation between u_i and v_i is zero, then the equations for w_i and $w^*_i - w_i$ form a simply recursive system.

---

72 A test of the null hypothesis that the system is recursive is described in Dhrymes (1986), "Limited Dependent Variables," cited above.

73 This description of Heckman's general approach follows Heckman (1974), "Shadow Prices, Market Wages," cited above, and Killingsworth (1983), "Second-generation studies of static models," cited above.

In general, however, u_i and v_i will be correlated since, when u_i and e_i are distributed independently with zero means,

$$Cov(u_i,v_i) = E(u_i v_i) = E[u_i(u_i - e_i)] = Var(u_i) \neq 0.$$

Only if $Var(u_i) = Cov(e_i u_i)$ will $Cov(u_i,v_i) = E[u_i(u_i - e_i)] = 0$. If persons with higher-than-average values of w_i have higher-than-average values of w^*_i, for example, and the correlation between u_i and e_i approaches the variance of e_i, it may be the case that $Cov(u_i,v_i) = 0$. However one generally expects that $Var(u_i) > Cov(u_i,v_i)$, in which case the possibility of selection bias should be addressed in estimation.

To appreciate the motivation for the estimation procedures that are employed to compensate for this form of selection bias, notice that when $w_i - w_i^* = 0$,

$$v_i = -[\, a' - a + D\, d_d - S\, b_s + C\, (d_c - b_c)].$$

Let F represent the cumulative distribution function for v_i. By substitution, the probability that person i will be employed is:

$$\begin{aligned} p_i &= PR\,(w_i > w_i^*) \\ &= PR\,(v_i < v^*) \\ &= F[a - a' - D\, d_d + S\, b_s - C\, (d_c - b_c)]. \end{aligned}$$

The corresponding probability distribution function for v_i is $f(v) = g(u,e)/g_e(u)$, where g is the joint probability distribution for u-e and g_e is the marginal distribution of u, given e.

This last equation might be estimated by setting $p_i=1$ for all employed persons and $p_i=0$ for non participants, and regressing p_i on C, D and S. But OLS procedures will not necessarily generate an estimate for p_i that lies within the (0,1) interval, as is required for a probability distribution function. Some econometricians adopt weighted least squares procedures to minimize this problem.[74] Theil seems to find weighted least-squares procedures to be acceptable when the cells for which weights are defined span relatively small intervals.[75] However other econometricians, including Pyndyck and Rotenberg, stress that greater explanatory power is obtained through the use of maximum likelihood procedures.[76]

Maximum likelihood, i.e., probit and logit procedures, involve estimating an equation of the form:

$$F^{-1}\,(p_i\,) = a' - a + D\, d_d - S\, b_s + C\, (d_c - b_c),$$

---

[74] Goldberger, for example, has recommended an approach in which weights are constructed from expected values of p_i, originally calculated with OLS; this approach is less costly than maximum likelihood estimation. However the costs of implementing maximum likelihood procedures have been reduced significantly since 1964. See Arthur Goldberger (1964), Econometric Theory, New York: Wiley, p. 250

[75] Henri Theil (1971), Principles of Econometrics, New York: Wiley, p. 633, and 1967, Sections 3.7-3.8

[76] Robert S. Pindyck and Daniel L. Rubinfeld (1981), Econometric Models and Economic Forecasts, Second Edition, New York: McGraw Hill, p. 310.

with nonlinear, iterative techniques.[77] The resulting parameter estimates are used to calculate $f(v_i)$, the conditional mean of u_i for the truncated sample of observations on persons who are employed. Notice that when the simplifying assumption that $Z=Z_i$ and $X=X_i$ is abandoned the conditional expected values vary across individuals, because v_i is a function of Z_i and X_i.

Consistent parameter estimates for the wage equation are subsequently obtained from the following augmented equation:

$$w_i = Z_i d + g \lambda_i ,$$

where the estimated value λ_i, a function of the predicted value for $F^{-1}(p_i)$, is entered as an additional explanatory variable in the augmented wage equation.

In the case of empirical models of male labor supply, procedures to control for selection bias often do not result in parameter estimates that are very different from those obtained by OLS procedures, according to Penceval.[78] But some procedure to test for selection bias is usually adopted for applications involving the labor supply of women, since the labor force participation rates of women remain much lower than those of men.

[77] This review of probit and logit estimation procedures follows Pindyck and Rubinfeld (1981), cited above, and Killingsworth (1983), "Second-generation studies of static models," cited above.

[78] John Penceval (1986), "Labor Supply of Men," cited above. A noteworthy exception to this statement is provided in the more recent work of Marshall and Zarkin (1987), cited above, which was published after Penceval's essay.

Appendix B: SIPP Estimates of Earnings and Wage Rates in 1984

Mean and Median Weekly Earnings

1984 SIPP Sample Estimates

median weekly earnings				
	full-time		part-time	
age	males	females	males	females
16to19	186.29	151.89	81.65	67.00
20to24	255.21	210.38	157.35	122.98
25to34	374.62	274.65	269.38	159.80
35to44	480.72	283.75	350.00	141.36
45to54	476.41	278.10	367.06	145.69
55to64	446.94	262.31	318.88	132.31
65+	312.57	201.81	95.72	75.83

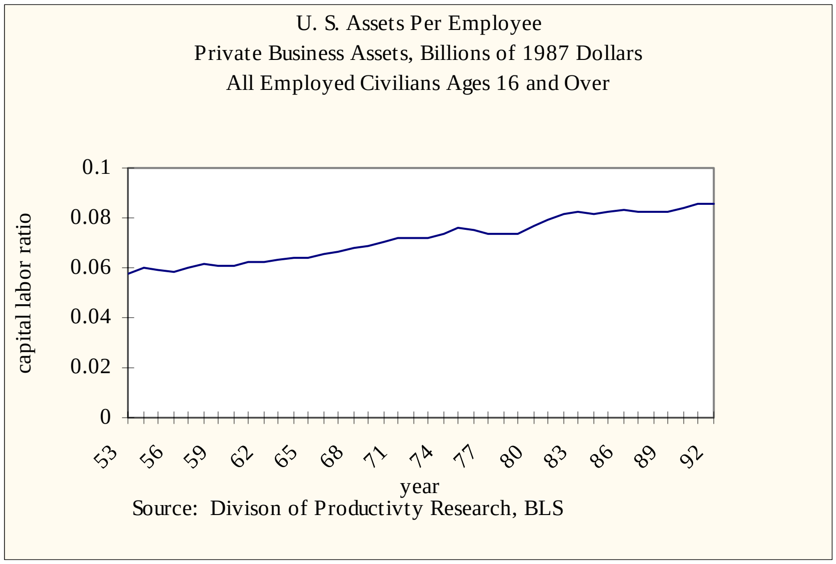
mean weekly earnings				
	full-time		part-time	
age	males	females	males	females
16to19	213.91	199.93	97.09	77.01
20to24	302.07	243.11	159.44	122.07
25to34	417.77	316.22	219.39	154.12
35to44	555.05	327.20	282.10	133.37
45to54	545.30	305.14	278.94	134.64
55to64	516.49	283.26	265.44	143.40
65+	392.70	223.83	149.70	87.47

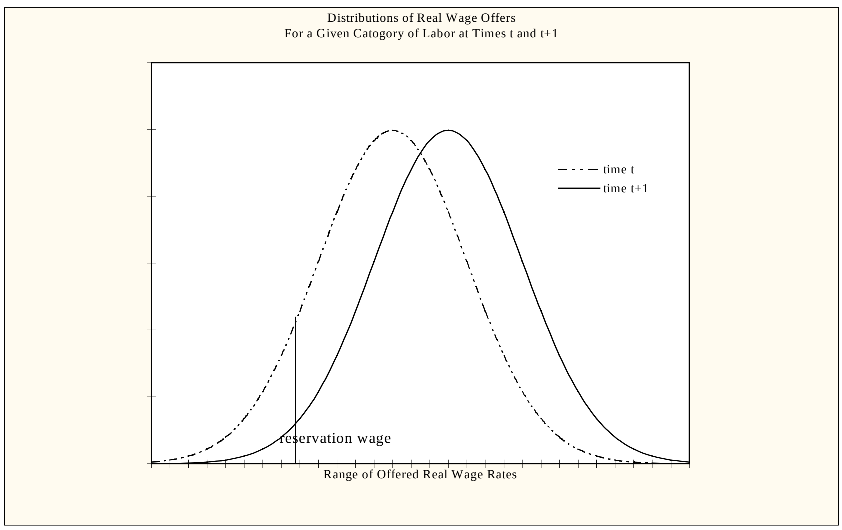
Mean Hourly Wage Rates
1984 SIPP Sample Estimate

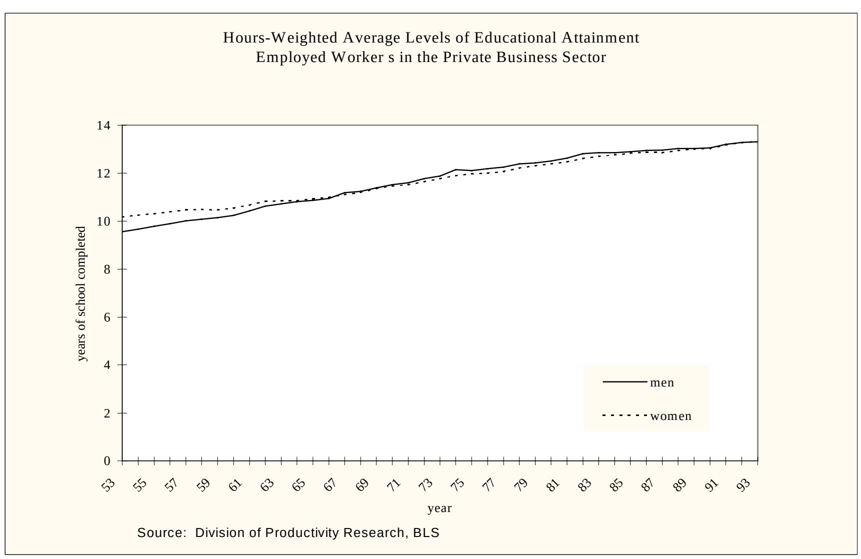
mean wage rates					
years of schooling	age	whites males	whites females	blacks males	blacks females
11	16to19	3.42	3.01	3.14	3.27
	20to24	5.05	3.84	3.43	3.56
	25to34	6.94	4.47	4.38	4.49
	35to44	8.95	4.81	6.94	4.03
	45to54	8.55	5.27	7.20	4.04
	55to64	9.50	5.07	7.61	4.83
	65+	5.87	4.28	**	**
12	16to19	3.88	3.58	4.41	3.43
	20to24	5.94	4.53	4.22	4.21
	25to34	8.07	5.74	6.25	5.19
	35to44	10.51	6.36	8.21	6.23
	45to54	10.80	6.17	8.32	5.61
	55to64	10.46	6.18	7.04	5.74
	65+	8.04	4.61	**	**
14	16to19	4.59	3.44	4.36	**
	20to24	5.38	4.72	5.06	5.44
	25to34	8.67	6.78	7.18	5.82
	35to44	11.56	7.10	9.75	8.71
	45to54	12.56	7.50	9.97	7.04
	55to64	11.61	7.53	10.79	7.05
	65+	10.38	6.65	**	**
16	20to24	6.73	6.09	**	5.87
	25to34	10.35	8.60	9.68	7.00
	35to44	14.25	8.10	11.69	8.26
	45to54	17.76	9.16	**	10.63
	55to64	18.03	8.92	**	**
	65+	9.57	5.88	**	**
17+	20to24	7.15	5.39	**	**
	25to34	10.87	8.95	10.37	6.76
	35to44	15.59	10.50	12.99	9.89
	45to54	16.97	9.40	11.04	12.13
	55to64	16.26	8.13	**	**
	65+	13.70	6.85	**	**

** identifies cells with fewer than 10 observations.
Note that SIPP sample sizes can be increased by pooling observations from adjacent panels that span the same calendar year. Work with pooled panels is planned for the next stage of the project.

Charts


U. S. Assets Per Employee
Private Business Assets, Billions of 1987 Dollars
All Employed Civilians Ages 16 and Over
capital labor ratio
0.1
0.08
0.06
0.04
0.02
0
53
56
59
62
65
68
71
74
77
80
83
86
89
92
year
Source: Divison of Productivty Research, BLS


Distributions of Real Wage Offers
For a Given Catogory of Labor at Times t and t+1
time t
time t+1
reservation wage
Range of Offered Real Wage Rates


Hours-Weighted Average Levels of Educational Attainment
Employed Worker s in the Private Business Sector
years of school completed
14
12
10
8
6
4
2
0
men
women
53
55
57
59
61
63
65
67
69
71
73
75
77
79
81
83
85
87
89
91
93
year
Source: Division of Productivity Research, BLS

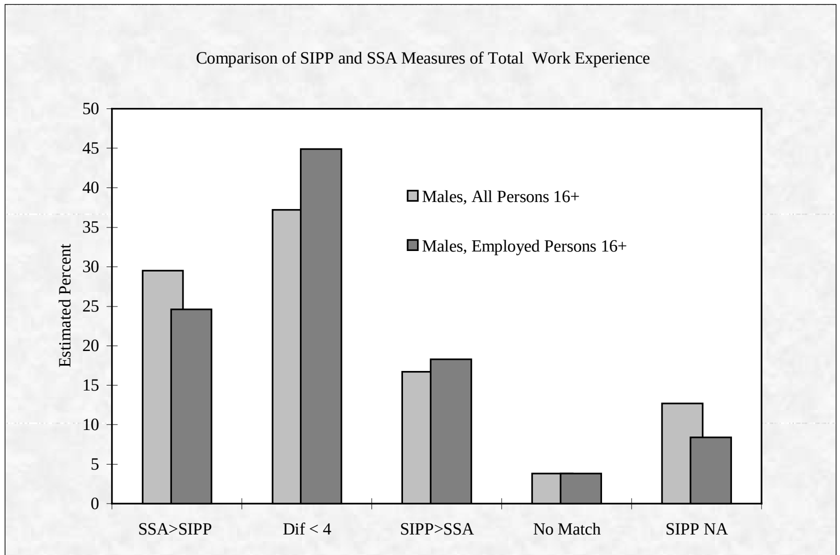
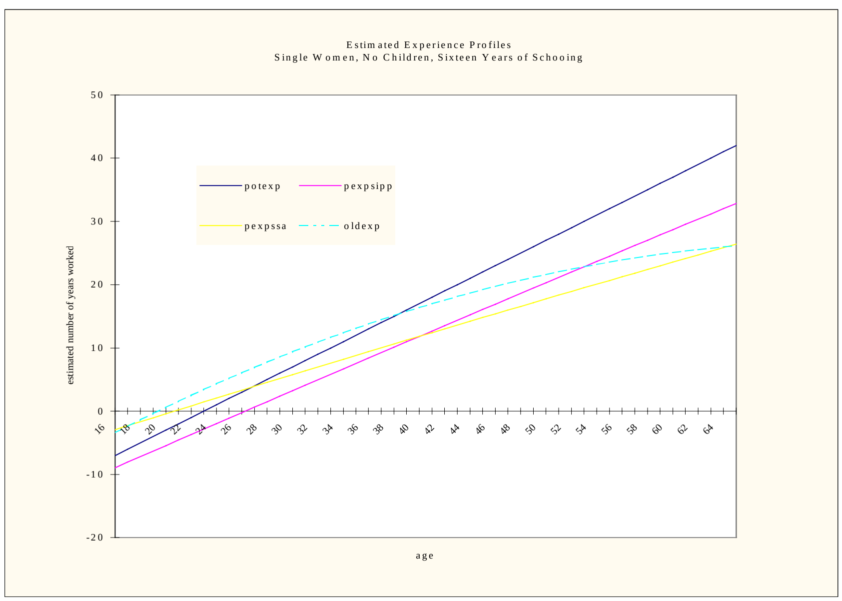
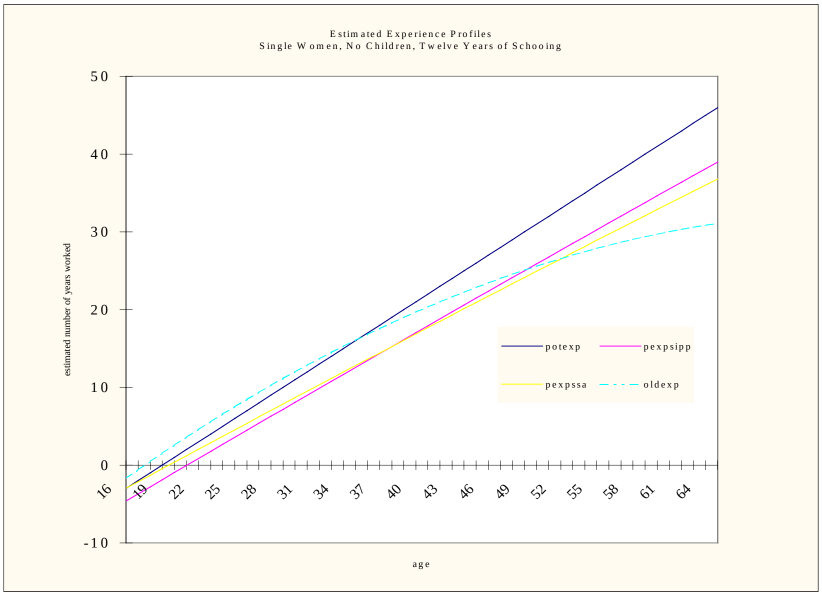
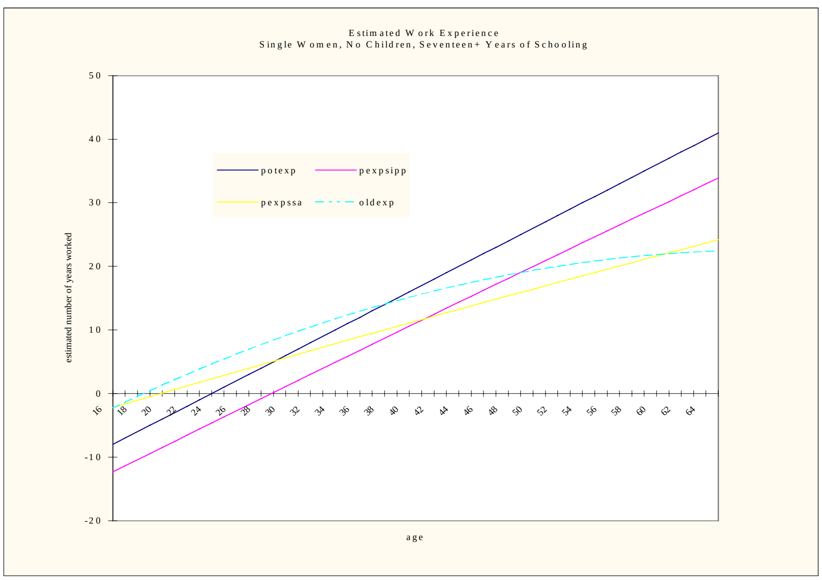
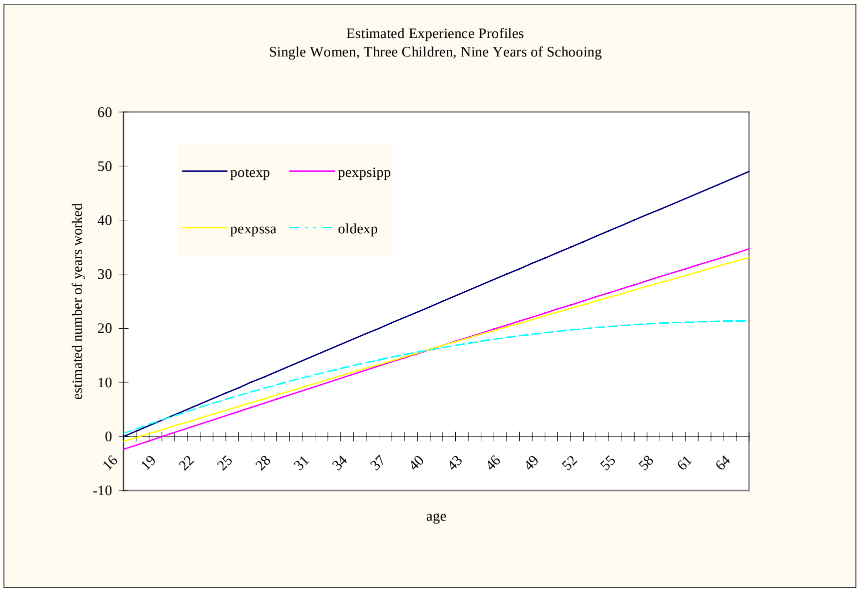

Comparison of SIPP and SSA Measures of Total Work Experience
Estimated Percent
50
45
40
35
30
25
20
15
10
5
0
Males, All Persons 16+
Males, Employed Persons 16+
SSA>SIPP
Dif < 4
SIPP>SSA
No Match
SIPP NA

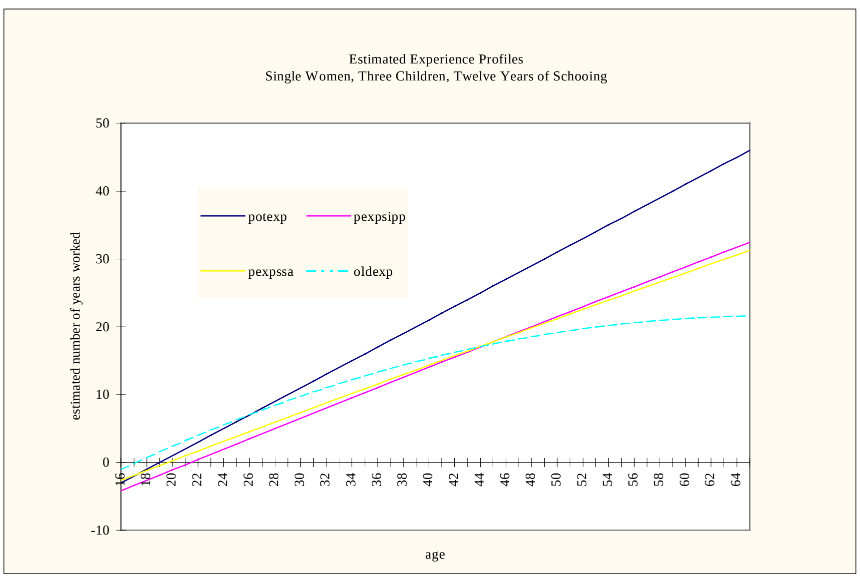
Chart 5

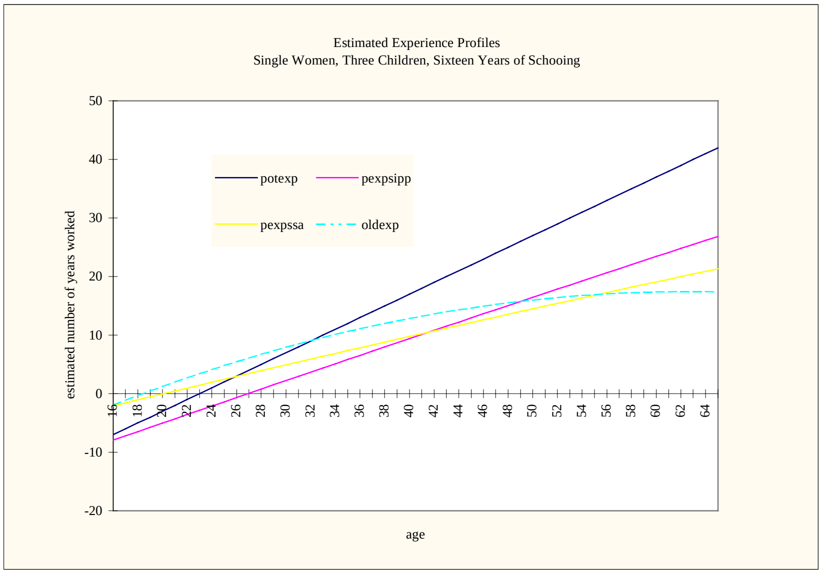

Estimated Experience Profiles
Single Women, No Children, Sixteen Years of Schooing
estimated number of years worked
50
40
30
20
10
0
-10
-20
potexp
pexpsipp
pexpssa
oldexp
age


Estimated Experience Profiles
Single Women, No Children, Twelve Years of Schooing
estimated number of years worked
50
40
30
20
10
0
-10
16
19
22
25
28
31
34
37
40
43
46
49
52
55
58
61
64
age
potexp
pexpsipp
pexpssa
oldexp

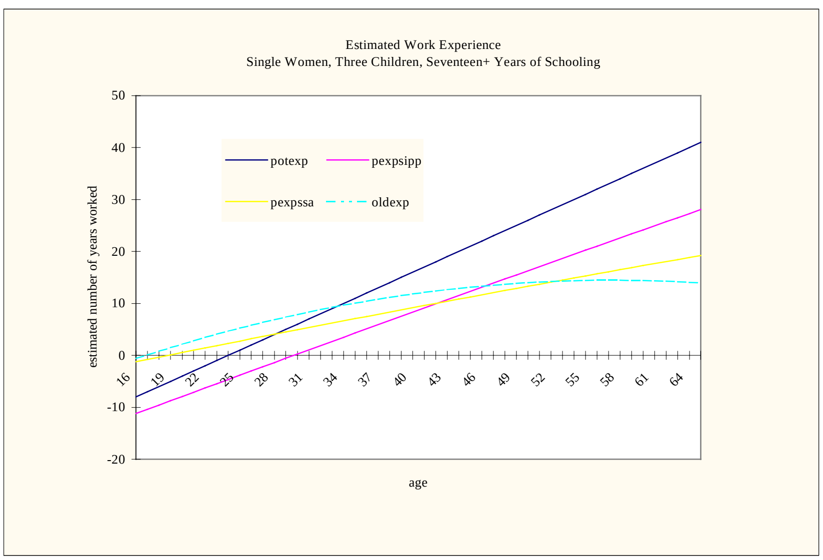

Estimated Work Experience
Single Women, No Children, Seventeen+ Years of Schooling
estimated number of years worked
50
40
30
20
10
0
-10
-20
age
potexp
pexpsipp
pexpssa
oldexp


Estimated Experience Profiles
Single Women, Three Children, Nine Years of Schooing
estimated number of years worked
age
potexp
pexpsipp
pexpssa
oldexp
60
50
40
30
20
10
0
-10
16
19
22
25
28
31
34
37
40
43
46
49
52
55
58
61
64

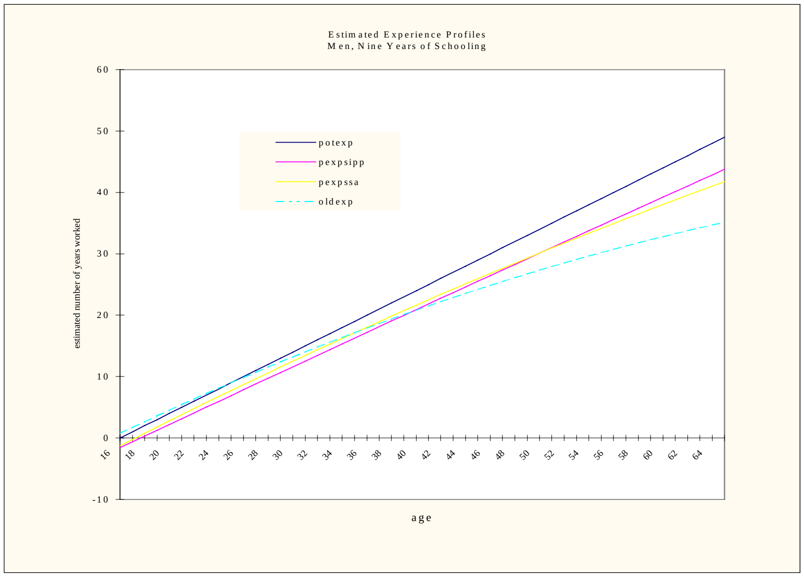

Estimated Experience Profiles
Single Women, Three Children, Twelve Years of Schooing
estimated number of years worked
age
potexp
pexpsipp
pexpssa
oldexp
50
40
30
20
10
0
-10
16
18
20
22
24
26
28
30
32
34
36
38
40
42
44
46
48
50
52
54
56
58
60
62
64

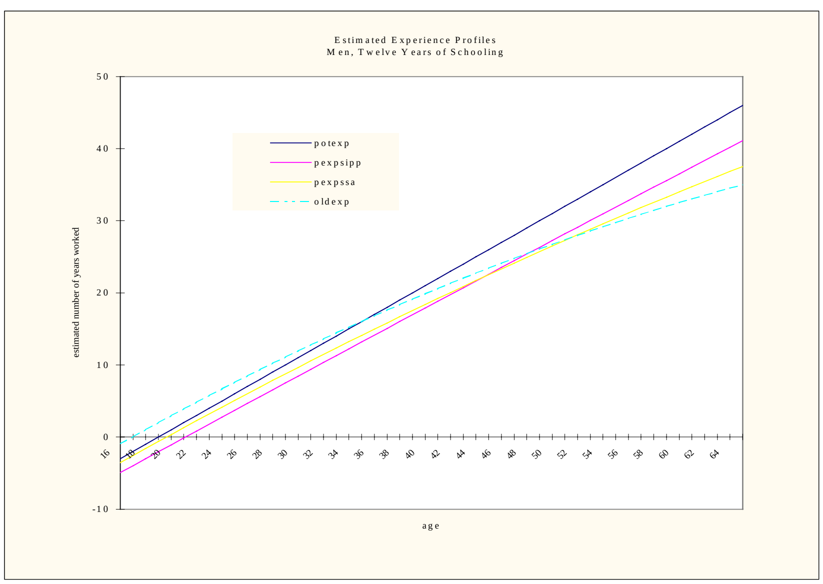


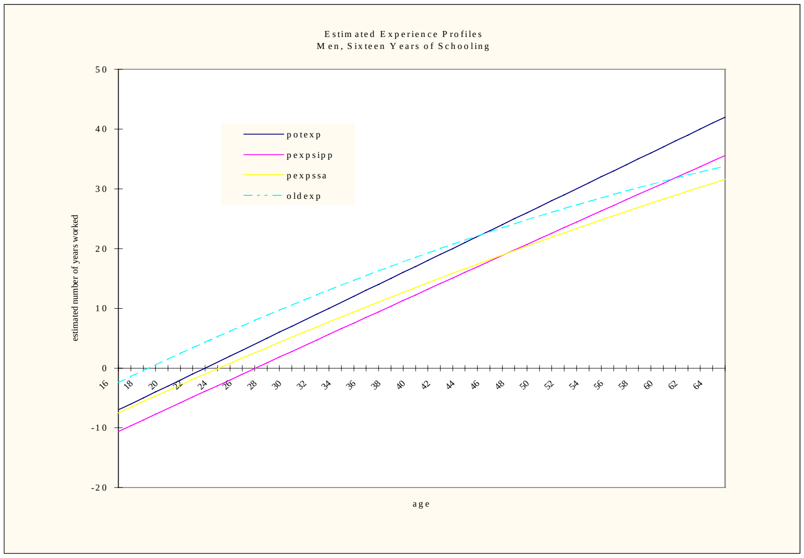
Estimated Experience Profiles
Single Women, Three Children, Sixteen Years of Schooling
estimated number of years worked
age
potexp
pexpsipp
pexpssa
oldexp

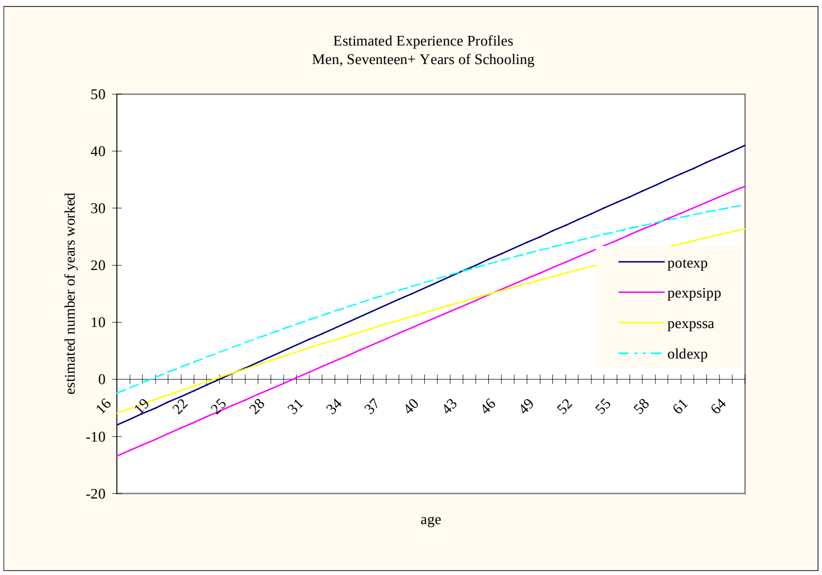


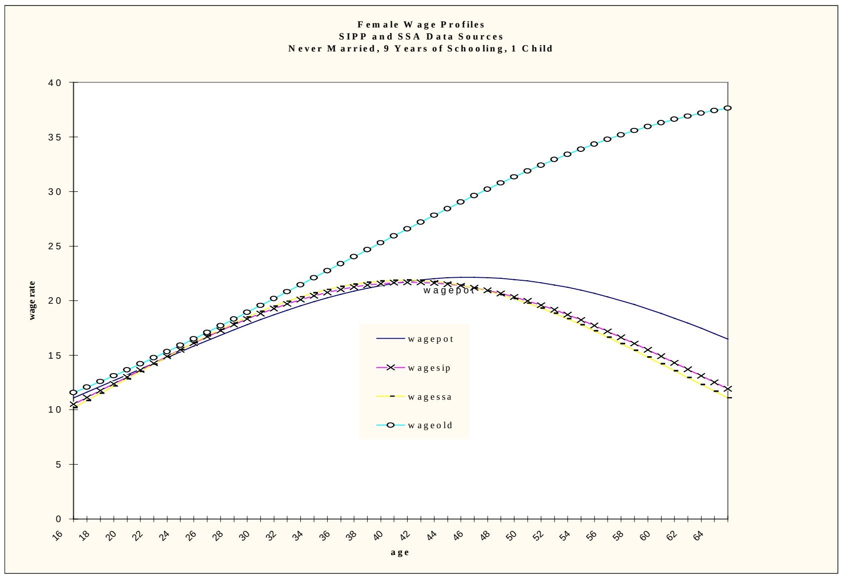
Estimated Work Experience
Single Women, Three Children, Seventeen+ Years of Schooling
estimated number of years worked
age
potexp
pexpsipp
pexpssa
oldexp

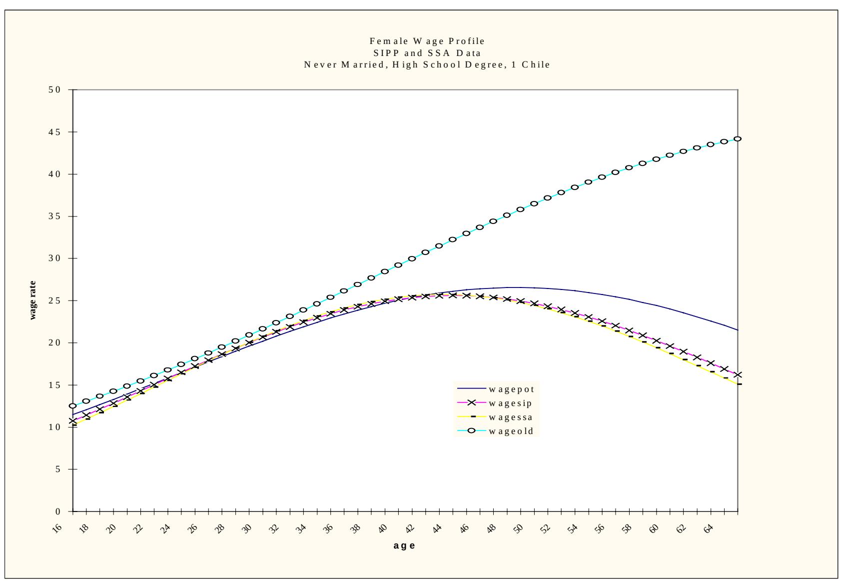

Estimated Experience Profiles
Men, Nine Years of Schooling
potexp
pexpsipp
pexpssa
oldexp
estimated number of years worked
age

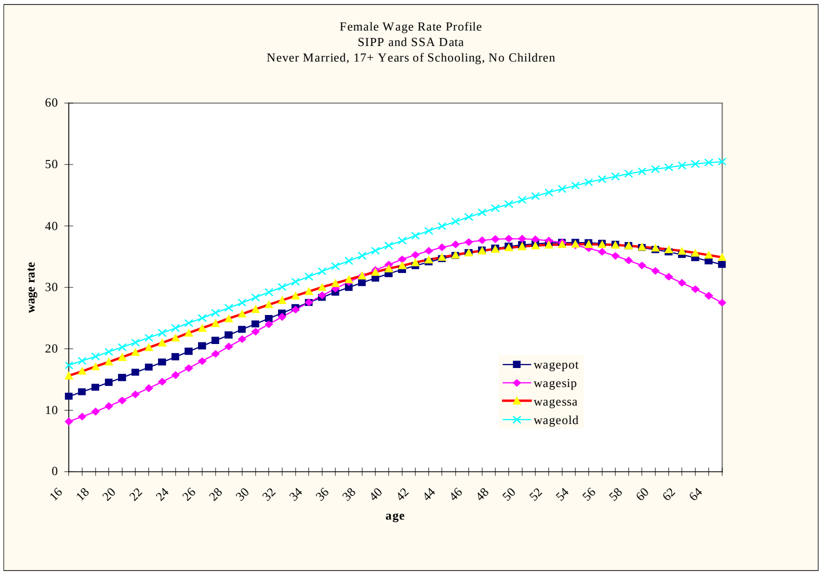

Estimated Experience Profiles
Men, Twelve Years of Schooling
potexp
pexpsipp
pexpssa
oldexp
estimated number of years worked
age

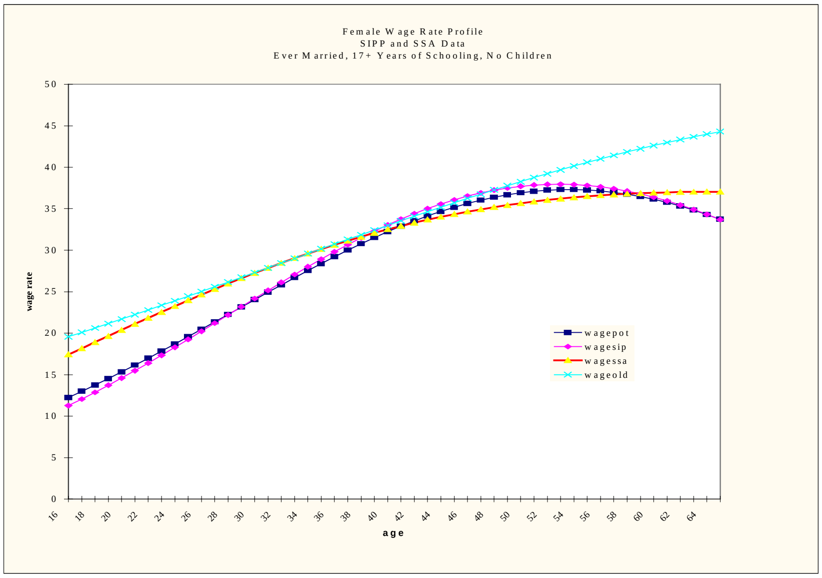

Estimated Experience Profiles
Men, Sixteen Years of Schooling
estimated number of years worked
age
potexp
pexpsipp
pexpssa
oldexp

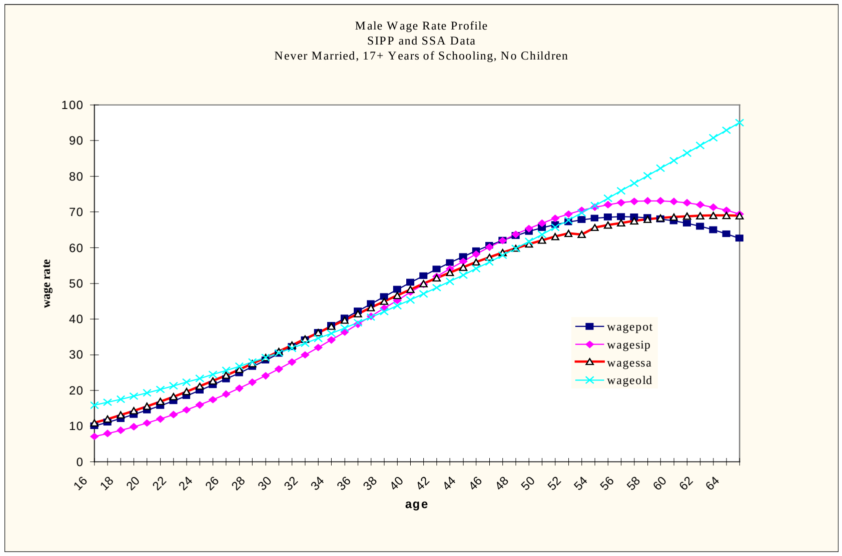

Estimated Experience Profiles
Men, Seventeen+ Years of Schooling
estimated number of years worked
age
potexp
pexpsipp
pexpssa
oldexp

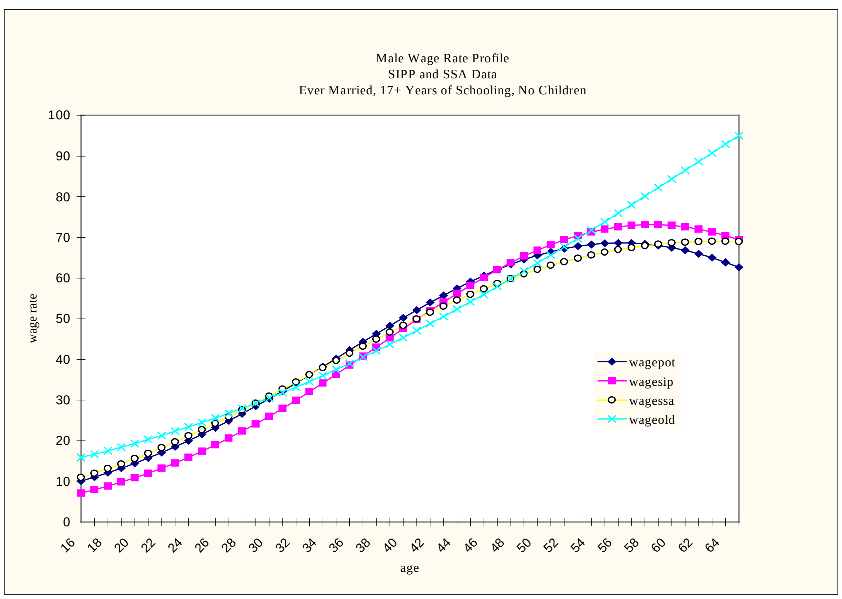

Female Wage Profiles
SIPP and SSA Data Sources
Never Married, 9 Years of Schooling, 1 Child
wage rate
age
wagepot
wagesip
wagessa
wageold


Female Wage Profile
SIPP and SSA Data
Never Married, High School Degree, 1 Chile
wage rate
age
wagepot
wagesip
wagessa
wageold


Female Wage Rate Profile
SIPP and SSA Data
Never Married, 17+ Years of Schooling, No Children
wage rate
age
wagepot
wagesip
wagessa
wageold


Female Wage Rate Profile
SIPP and SSA Data
Ever Married, 17+ Years of Schooling, No Children
wage rate
age
wagepot
wagesip
wagessa
wageold


Male Wage Rate Profile
SIPP and SSA Data
Never Married, 17+ Years of Schooling, No Children
wage rate
age
wagepot
wagesip
wagessa
wageold


Male Wage Rate Profile
SIPP and SSA Data
Ever Married, 17+ Years of Schooling, No Children
wage rate
age
wagepot
wagesip
wagessa
wageold